



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 07:50:13 ; Search time 3099 Seconds
(without alignments)
10574 301 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1126
Sequence: 1 gcattattagtggagca.....aataaatgtgtcatgacct 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: GenBank

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_to.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_to.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rtd.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	853.2	75.8	1752	9	BC002642	BC002642 Homo sapi
2	845.4	75.1	1643	6	AR209918	AR209918 Sequence
3	845.4	75.1	1763	6	AX329664	AX329664 Sequence
4	845.4	75.1	1763	9	HUMCATS	M90696 Homo sapien
5	842.4	74.8	1255	9	HUMCATS	M86553 Human cath
6	842.4	74.8	1255	9	S93414	S93414 cathepsin S
7	669	59.4	1279	10	BC002125	BC002125 Mus muscu
8	667.4	59.3	1305	10	AF038546	AF038546 Mus muscu
9	665	59.1	1296	10	MMCATHEPS	AF002386 Mus muscu
10	631	56.0	1330	10	RATCTHPS	L03201 Rattus norv
11	551.4	49.0	1234	4	HICATPS	Xb2001 H.taurus mk
12	551.4	49.0	1235	4	BOVCATS	M95211 Bovine cath
13	320	28.4	1212	5	CYICYP	L30111 Cyprinus ca
14	313.4	27.8	491	10	MMU18466	Y18466 Mus musculu
15	313.2	27.8	934	4	AF292030	AF292030 Sus scro
16	310	27.5	990	6	AB7768	AB7768 Sequence 2
17	310	27.5	1482	6	AR000416	AR000416 Sequence
18	310	27.5	1482	9	S79895	S79895 OC2-catheps
19	310	27.5	1619	6	AR208697	AR208697 Sequence
20	310	27.5	1619	6	I19038	I19038 Sequence 1
21	310	27.5	1661	9	HS013665	U13665 Human cath
22	310	27.5	1669	6	AR109588	AR109588 Sequence
23	310	27.5	1669	6	AR194312	AR194312 Sequence
24	310	27.5	1669	9	HSOC2RNA	X82153 H.sapiens m
25	310	27.5	1699	9	BC016058	BC016058 Homo sapi
26	308.4	27.4	990	6	AR109592	AR109592 Sequence
27	308.4	27.4	990	6	AR194316	AR194316 Sequence
28	308.4	27.4	990	9	AF070927	AF070927 Macaca fa
29	308.4	27.4	1029	9	AF124092	AF124092 Macaca mu
30	306.8	27.2	990	6	AR109593	AR109593 Sequence
31	306.8	27.2	990	6	AR194317	AR194317 Sequence
32	306.8	27.2	1597	9	HS010280	U20280 Human cath
33	303.8	27.0	1599	4	PABOC2	D14036 Rabbit mRNa
34	300.2	26.7	990	6	AR072914	AR072914 Sequence
35	300.2	26.7	990	6	E28867	E28867 Polynucleot
36	300.2	26.7	990	6	E51957	E51957 Polynucleot
37	300.2	26.7	1248	10	AF010306	AF010306 Rattus no
38	292.4	26.0	990	6	AX305997	AX305997 Sequence
39	292.4	26.0	990	10	MMPPCATHK	X94444 M.musculus
40	289.4	25.7	1044	6	AX063689	AX063689 Sequence
41	262.8	23.3	1420	5	AF358668	AF358668 Oncorhyn
42	256	22.7	1259	3	AF347207	AF147207 Artemia f
43	234.6	21.3	1098	3	AF320565	AF320565 Rhodnius
44	239.2	21.2	1178	9	AF070448	AF070448 Homo sapi
45	239.2	21.2	1342	6	E58936	E58936 Novel human

ALIGNMENTS

RESULT 1
BC002642
LOCUS Homo sapiens, cathepsin S, clone MCC:3886 IMAGE:3610589, mRNA,
DEFINITION complete cds.
ACCESSION BC002642
VERSION BC002642 1 GI-12803614
KEYWORDS MCC
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1752)
AUTHORS Strausberg,R.
TITLE Direct Submission

JOURNAL

Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Hakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAL Plate: 12 Row: C Column: 21

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 806607.

FEATURES

Location/Qualifiers
1..1752

/organism="Homo sapiens"

/db_xref="locusid:1520"

/db_xref="taxon:9606"

/clone="MGC:3886 IMAGE:3610589"

/tissue_type="Pancreas, adenocarcinoma"

/clone_lib="NIH_MGC_39"

/lab_host="DH10B-K"

/note="Vector: pOTB"

72..1067

/codon_start=1

/product="cathepsin S"

/protein_id="AAH02642.1"

/db_xref="GI:12803615"

/translation="MKRLVCLVLCSSVAQLKDPDLDDHHHLKKTKYKQYKKEKE
EVRRLLEWRNLFVLDHSHGMSHISYDGLNHLGMDTSEVMSLSLRVPSQWQ
RNITYKNWLPDSDVREKGVTEVYKQSCGAWFAVSALEALQKDKTKLY
SLSQNLSKYTEKYGKNGCGFMFTAFQYIIDNKIGDSASYPYKAMDRCQYDSK
YRAATCSKYTELIPYGRDVLKFAVANKPVSVCVDPARHPSFLYRSYVYEPSCQNV
NHGVLVVGYGDLNCKEYVLVKNWGHNFGEQYIRMARKNHNHGGIASFSPSPPEI"

BASE COUNT
ORIGIN

Query Match 75.8%; Score 853.2; DB 9; Length 1752;
Best local Similarity 87.1%; Pred. No. 4.7e-211;
Matches 984; Conservative 0; Mismatches 138; Indels 8; Gaps 4;

QY 1 GCATATATAGTATGGGAGCCTGCAGGTTCTTTC---ATGAATGGCTAGTTGGCTGC 57
DB 34 GGACTCTTACGCTGGGAGCACTGCTGGTCTATCAACAATGAACGGCTGGTTGTGTC 93
QY 58 TTCCGTTCTGCTCTATGCTGGCAGCACTACATAAAGATCCCACTCTGGATCATCACT 117
DB 94 TCTTGGTCTGCTCTTCTGAGTGCATAGTTGATATAAATATATATCTTGGATCAGCACT 153
QY 118 GGAATCTCTGGAAGAAACCTACAGCAAAAT---CAAGAGAGAGATGAGGAAGTAGCAC 176
DB 154 GGCATCTCTGGAAGAAACCTATGCGCAACAACTACAGGAAAGAAATGAAGAAGCAGTAC 213
QY 177 GGGCTCTCATCTGGAAAAAATCTAAATTTTGTGATGCTTCAATCTTGGAACTCAA 236
DB 214 GAGCTCTCATCTGGAAAAAATCTAAAGTTTGTGATGCTTCAATCTTGGAACTCAA 273
QY 237 TCGGAATGCATTTCATATCATCTAGGCAATGAACCATCTGGCAGACATCACTGGTGAAGAAG 296

DB 274 TGGGAATGCACCTATACCATCTGGGATGAACCACTGGGAGACATGACCACTGAAGAAG 333
QY 297 TGATATCTTTGATGGTTCCCTGAGAGTTTCAAGAGCAATGGAGAGAAATGTCTATTATA 456
DB 334 TGATCTCTTTGATGAGTTCCCTGAGAGTTCCCAAGCCAGTCCGAGAGAAATATACATA 493
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATTCCTGACACTGGAGAGAGAGGCTGTGTTA 416
DB 394 AGTCAAAAGCTTAATGATATATGCTGATTCCTGAGATTCGAGAGAGAAAGGTTGTGTA 453
QY 417 CTGAAGTGAATACCAAGGTTCTTGTGCTGCTTGTGGGCTTTCACGGCTTGGGGGGCC 476
DB 454 CTGAAGTGAATATCAAGGTTCTTGTGCTGCTGCTGAGGCTTTCAGTGTCTGGGGGCC 513
QY 477 TGAAGCACAACTCAAGCTAAAACAGGAAAGCTGCTGCTGATGATGATGATGATGATG 536
DB 514 TGAAGCACAGCTCAAGCTGAAAACAGGAAAGCTGCTGCTGATGATGATGATGATGATG 573
QY 537 TAGATTCTCAACTGAAAATATGGAATTAAGGCTGCAATGGGCTTTCATGATCACTG 596
DB 574 TGAATTGCTCACTGAAAATATGGAATTAAGGCTGCAATGGGCTTTCATGATCACTG 633
QY 597 CTTTCCAGTATATTATTAACAACGGCAATGATTTACAGAGCTTCTCTATCTTCTTCTA 656
DB 634 CTTTCCAGTATATTATTAACAACGGCAATGATTTACAGAGCTTCTCTATCTTCTTCTA 693
QY 657 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716
DB 694 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 753
QY 717 CTGAACCTTCCCTTGGCAGTCAAGATGCCCTTAAAGAGCTGTGGCAATTAAGGAGCT 776
DB 754 CTGAACCTTCCCTTGGCAGTCAAGATGCCCTTAAAGAGCTGTGGCAATTAAGGAGCT 813
QY 777 TGCTGTGGCTATAGATGGCAGCCTATCTTCTTCTCTCTACAGAGCTGCTGCTACT 836
DB 814 TGCTGTGGCTATAGATGGCAGCCTATCTTCTTCTCTCTACAGAGCTGCTGCTACT 873
QY 837 ATGAACCATCTCTATCTCAAGATGTAATCTATGATGATGATGATGATGATGATGATG 896
DB 874 ATGAACCATCTCTATCTCAAGATGTAATCTATGATGATGATGATGATGATGATGATG 933
QY 897 TTAATGGGAAGACTACTGCTGCTGTAAGAGCTGGGGCTTCACTTGGTGAAGCAAG 956
DB 934 TTAATGGGAAGACTACTGCTGCTGTAAGAGCTGGGGCTTCACTTGGTGAAGCAAG 993
QY 957 GATATATTCGATGGGCAAGAAACAGTGGAAATCACTTGGGATTTGCTGATTTCTTCT 1016
DB 994 GATATATTCGATGGGCAAGAAATCACTTGGGATTTGCTGATTTCTTCTTCTTCTT 1053
QY 1017 ACCGAGAAATCTA---GACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1073
DB 1054 ACCGAGAAATCTTACGAGCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1113
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAATAAATGATGATGATGATGATGATG 1122
DB 1114 TCTTAATTTAATTTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1163

RESULT 2

AP209918

LOCUS AP209918 1643 bp DNA linear PAT 20 JUN-2002

DEFINITION Sequence 1 from patent US 6487629.

ACCESSION AR209918

VERSION AR209918.1 GI:21512013

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1643)

AUTHORS Schneider, P., Yamamoto, K.K. and French, C.K.

TITLE Use of cathepsin S in the diagnosis and treatment of endometriosis


```
/db_xref="taxon:10090"
/tissue_type="spleen"
100..1092
/codon_start=1
/product="cathepsin S"
/protein_id="CAA05360.1"
/db_xref="GI:385078"
/translation="MAVLGAPGVLCNGATAPRTLDHWDLWKKTHEKEYKDKNEE
VRLIWEKNLKFIMHNLDESMTGYVGMNDMDTNEEILCRMGALRIPIQSPKT
VTFRSYNRPLPTDVRKRGCTEVYKQSGCAGWAFSAVGALEQLKTKGLISL
SAQNLVDCSTBEKYNKGGGPMTEAFQYIIDNGIEADASYPYKAMDEKCHYNSKN
RAATCSRYIOLPDDDEDALKEAVATKPVSVGIDASHSFYFKSGVYDDPSCITGVN
HCVLVGVGILDGKDYWLKNSWGLNFGDQGYIRMARNNKHGIIASYCSYPEI"
100..150
sig_peptide
mat_peptide
436..1089
/product="cathepsin S"
BASE COUNT 362 a 274 c 323 g 337 t
ORIGIN
Query Match 59.1%; Score 665; DB 10; Length 1296;
Best Local Similarity 77.3%; Pred. No. 4.2e-162;
Matches 874; Conservative 0; Mismatches 245; Indels 12; Gaps 5;
QY 2 CATTTATTAGTAGGAGCACCCTGCAGGTTCTTTCATCAAAATGGCTAGTGGGCTGTCC 61
  || || || || || || || || || || || || || || || || || || || || ||
Db 61 CAGGATGACCATGAGGGCTCTCGCCACGCTGCCATAGATGGCTGTTTGGATGCCCT 120
QY 64 GTTGTCCTCTATGTCAGTGGACAAAGTACATAAAGATCCCACTGTCATCACTGGAA 121
  || || || || || || || || || || || || || || || || || || || || ||
Db 121 GGTGTGCTCTGTCGGCATTGGAGCAACTGCAGAGAGA-CCTACCCCTGGACCACCTGGGA 179
QY 122 TCTCTGGAGAAACCTACAGCAAAACAAAT-CAAGGAAGAGAATGAGCAAGTAGCAGCGG 180
  || || || || || || || || || || || || || || || || || || || || ||
Db 180 TCTCTGGAGAAACCTACAGCAAAACAAATGAGATGAGATGAGATGAGATGAGATGAG 239
QY 181 TCTCATCTGGAGAAACCTAAATTTGATGCTTCAACAATCTGGAACACTCAATGGG 240
  || || || || || || || || || || || || || || || || || || || || ||
Db 240 TCTCATCTGGAGAAACCTAAATTTGATGCTTCAACAATCTGGAACACTCAATGGG 299
QY 241 AATGATTCATATGATCTAGGCTGAGCAATCTGGGAGACATGCTGGTGAAGAAGTGAT 300
  || || || || || || || || || || || || || || || || || || || || ||
Db 300 AATGATTCATATGATCTAGGCTGAGCAATCTGGGAGACATGCTGGTGAAGAAGTGAT 359
QY 301 ATCTTTTGATGGGTTCCCTGAGAGTTCCAGCAATGGCAGAGAAATCTCACTATAGGTC 360
  || || || || || || || || || || || || || || || || || || || || ||
Db 360 GGTGCGATGGGTTCTGAGAAATTCGCGAGAGTCTCCGAGAGTCTGTCATTTTCAGGTC 419
QY 361 AAACCTCAATCAGAAATTCGCTGATTCGAGCTGGAGAGAGAGGCTGTGTTACTGA 420
  || || || || || || || || || || || || || || || || || || || || ||
Db 420 ATACTCTAATCGGACATTCGCTGACATGCTGGAGAGAGAGGCTGTGTTACTGA 479
QY 421 AGTGAATACAGGGTTCTTGTGTGTTTGTGAGCTTTCAGGCTGTGGGGGCTTGA 480
  || || || || || || || || || || || || || || || || || || || || ||
Db 480 GGTGAATACAGGGTTCTTGTGTGTTTGTGAGCTTTCAGGCTGTGGGGGCTTGA 539
QY 481 ACACAACTGAGAGCTAAACACAGGAAGAACTGGTCTCTGAGTCCAGAGAACTGGTAGA 540
  || || || || || || || || || || || || || || || || || || || || ||
Db 540 AGGGAGCTGGAAGCTGGAACAGGGGAAGCTGATATCCCTCAGTCTCAGAACTGGTGA 599
QY 541 TTGCTCAACT---GAAATAATPGGAATAAAGGCTGCAATGGCGGTTTCATGACAACTGC 597
  || || || || || || || || || || || || || || || || || || || || ||
Db 600 CTGCTCAATGAGAAAGAACTATGGGAATAAAGGCTGTGGAGCGGCTACATGACCGAAGC 659
QY 598 TTTCAGTATATTATTGATTAACAGGGGATTGATTCAGAGCTTCCTATCCCTCAAAAGC 657
  || || || || || || || || || || || || || || || || || || || || ||
Db 660 TTTCAGTATATTATTGATTAACAGGGGATTGATTCAGAGCTTCCTATCCCTCAAAAGC 719
QY 658 CATGAATGAAAGCTGACATATGATCAAAAAGAGGAGCTGCCACATGTTCAAGATATAC 717
  || || || || || || || || || || || || || || || || || || || || ||
Db 720 CATGATGAAAGCTGACATATTAACATAAAAAGAGGAGCTGCTAGTGTTCAGAGTACAT 779
QY 718 TCAACTTCCTCTTTGGCAGTGAAGATGCTTAAACAAAGCTGTGGCCAAATAAAGACCTGT 777
  || || || || || || || || || || || || || || || || || || || || ||
```


KEYWORDS cysteine proteinase.
SOURCE Cyprinus carpio female adult ovary cDNA to mRNA.
ORGANISM Cyprinus carpio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Cyprinus.
REFERENCE 1 (bases 1 to 1212)
AUTHORS Tsai, Y.-J. and Huang, F.-L.
TITLE Molecular cloning of a cysteine protease from carp ovary
JOURNAL Unpublished (1994)
FEATURES
source location/Qualifiers
1..1212
/organism="Cyprinus carpio"
/db_xref="taxon:7962"
/sex="female"
/tissue_type="ovary"
/dev_stage="adult"
53..1048
/note="putative"
/codon_start=1
/product="cysteine proteinase"
/protein_id="AAA49207.1"
/db_xref="GI:463046"
/translation="MHGSLLPVCCSAALAFNTNLDHMLWKKTHNKYSSKDEE
LGRRLERNGLITLHLELDLHLSYDLGNHMGDMTIEILQLTATIRVPGFKR
QTAIEVFGSGAAIPDLMDREKGYVSVKQACGSCWAFSGVLEGLMKTKLV
DLSPNILVPCSSYGNCGGLMSAAFOYIDNGGIDSESSYPYGVGOCRNPSQ
LAANTKYIYVQGDDEALQAVANI GPI SVAIDATHPOFILYRSYVNDPSCNTIN
HAALVAGYCAIAGQDFWLVKNSWGTGFGVGGYIRNARNQNNMCGIAHTPSIRLGN"
polyA_site 1212
BASE COUNT 314 a 256 c 313 g 329 t
ORIGIN
Query Match 28.4%; Score 320; DB 5; Length 1212;
Best local Similarity 61.4%; Pred No. 2.2e-72;
Matches 567; Conservative 0; Mismatches 350; Indels 7; Gaps 3;
QY 100 CCACTGTGATCATCACTGGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAG 158
DB 114 CAAATCTAGACCACTGGGAGTGTGGAAGAACGCAATCAACAGTTTACTCCAGCA 173
QY 159 AGAATGAGCAATAGTAAAGGCTCTATCTGGAAGAAACCTAAATTTGTGCTTC 218
DB 174 AGATGAAAGAACTGGGAGAGAGATGTTGTGGAAGAAACCTGGGCTTATCACTTC 233
QY 219 ACAATCTGGAACACTCAATGGGAATGCAATCATATCATATGAGCAATCAACCATCTGGGAG 278
DB 234 ACAACCTGAGACCTCTATGGGCTGATTCATATGAGCTGGGCAATGAACCATCTGGGTG 293
QY 279 ACATGACTGTGAAGAAGTGATATTTTATAGTTTCTGAGAGT---TCCAGCCCAAT 335
DB 294 ACATGACAAACAGGAGATCTCCAAAGCTTAGCCACGATTCCTGCTCCTCTCTGCTTTA 353
QY 336 GGCAGAGAAATGTCATATAGTCCAAATCTCAATCAGAAATAGTGAATCTGGGACT 395
DB 354 AGNGGCAACACAGAGTTTGTGGGCTCATCTGGAGCTGCTATCCAGACTCTGGACT 413
QY 396 GGAGACAGAGGGCTGTCTTACTGAGCAATAACACGGTCTTGTGGTCTGTGGG 455
DB 414 GGAGACAGAGGGATATGCTCTGAGTGTGAAGACAGGGGTGCTGCTCATGCTGGG 473
QY 456 CTTTACGGCTGTGGGGCTCTGGAGGACACTCAAGCTAAACAGCAAGCTGGGT 515
DB 474 CGTTTACGCTGTGGGCTCTTGAAGGTGAGCTGATGAACACCACTGGAAGCTGGCG 533
QY 516 CTCTGAGTGACAGAACTGTAGATTGCTCAACTGAAAAATATGGGAATAAAGGCTGCA 575
DB 534 ACCTCAGTCTCAGAACTGAGTGGAGTGTTCCTGCT---CAGTTACGAGCAATACGAGCTG 590
QY 576 ATGGCGGTTTCATGACAACTGCTTTCAGTATATATTGATTAACACCGCATGATTCAG 635
DB 591 GTGGGGTTGATGAGTGTGCTCTCCAGTATGTTTATGATAATGTTGGAATAGACTCCG 650

QY 636 AAGCTTCTATCCCTCAAAAGCATATGATGAAAGATGATATGATCAATCAAAAGAGAG 695
DB 651 AGTCATCTTACCCCTTATCAAGGAGTCAAGGGCAGTCAGATCAATCCCATCTCAGTGG 710
QY 696 CTGCCACATCTTCAAGCTATCACTCAACTTCCCTTTGGCAGTGAAGATGCCCTAAAAGCAAG 755
DB 711 CAGCAAACTGCACCACTACTATTACCTCTCTAGGAGATGAGAGAGCTTAAATAGG 770
QY 756 CTGTGGCCAATAAAGGACCTGTCTGTGGCTATAGATGGAGCACTATCTTTCTCTCT 815
DB 771 CTGTGTCAACATCGGGCTATTTCACTTGCCTATGATGACACCCCTCAGTTATCC 840
QY 816 TGTAAAGAGAGTGTCTATATCAATCAATATATATATATATATATATATATATATATAT 875
DB 831 TGTACCCAGCTGGAGCTTACAAATGATCACTCTGACCTCAATCAATCAATCAATCAAT 890
QY 876 TAGTGGTGGCTATGTTAACTTAAATGGGAAGCACTACTGCTTTGGTGAAGAAATAGTGGG 945
DB 891 TGGCTGTGAGATAAGTGTGATTGTGTAAGAAATTTTGTGTTGTAAGAAATAGTGGG 950
QY 936 GCCTCAACTTTGGTGACCAAGGATATATGGAAGGCAAGCAAGCAAGCAAGCAAGCAAG 945
DB 951 GTACCGGATTTGGAGATGGTGGCTACATCCCTATGCGCCAGAAACACAGAAACATGCTG 1010
QY 996 GGATTGCTAGTTATCTCTCTTAC 1019
DB 1011 GCATCGCTCATACCCCTCTATCC 1034
RESULT 14
MMU18466
LOCUS Mus musculus mRNA for cathepsin S, partial. 491 bp mRNA linear Feb 17 AUG 1999
DEFINITION Y18466
ACCESSION Y18466
VERSION Y18466.1 GI:3929822
KEYWORDS cathepsin S.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 491)
AUTHORS Soderstrom, M., Salminen, H., Glumoff, V., Kirschke, H., Aro, H. and Vuorio, E.
TITLE Cathepsin expression during skeletal development
JOURNAL Biochim. Biophys. Acta 1446 (1-2), 35-46 (1999)
MEDLINE 99326135
PUBMED 10395917
REFERENCE 2 (bases 1 to 491)
AUTHORS Soderstrom, M.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-1998) M. Soderstrom, Dept. of Med. Biochemistry and Biology, University of Turku, Kiinamyllynkatu 10, 20520 Turku, FINLAND
FEATURES
source location/Qualifiers
1..491
/organism="Mus musculus"
/strain="C57 black"
/db_xref="taxon:10090"
/tissue_type="cartilage"
<1..>491
/codon_start=1
/product="cathepsin S"
/protein_id="CAA77184.1"
/db_xref="GI:3929823"
/db_xref="SPTREMBL:O70370"
/translation="CGSWAFSAVGALEFGIKTKFKLISLAUNLVKSNEEKYGNK
GGAGGYMTFAFVYIILNGSTADASPYKATPEKQVNSPKRAATCSRYIQLDPGDEL
ALKEAVATKPVSVGIDASISISFFFKSGVDDPSCTGNVNHGLVVGYGTLTKKLYW
LVK"
primer_bind 1..21
/note="PCR forward primer"
primer_bind 474..491
/note="PCR reverse primer"

[illegible]

Search completed: July 1, 2003, 10:20:28
Job time : 3122 secs

Tue Jul 1 14:40:54 2003

DNA chip analysis as given in the specification, and comparing the expression level to an expression of Gs in an unactivated GC, where differential expression (M2) GA by contacting GC with an agent. Also included are modulation of at least one gene in Gs: (2) screening (M3) that alters the expression of GA or an inflammation (especially for an agent capable of modulating GA or an inflammation, exposure of a chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection and M5 is parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1763 BP; 525 A; 345 C; 400 G; 493 T; 0 other;

Query Match 75.1%; Score 845.4; DB 24; Length 1763;
Host Local Similarity 87.2%; Pred. No. 7.3e-239;
Matches 966; Conservative 0; Mismatches 136; Indels 9; Gaps 5;

1 GCATTATTAGTATGAGGACCTGCAGGTTCTTTTC---ATGAATGGCTAGTTGGGCTGC 57
99 GGACTCTTACTGTAGGACCACTGCTGGTCTTATCAACAATGAACGGCTGTTGTGTGC 158
58 TTCCGCTGTCCTTATGCAGTGGCACAAGTACATAAAGATCCCACTCTGCATCACT 117
159 TCTTGGTGCTCTCTGCAGTGGCAGACATGATGATGATGATGATGATGATGATGAT 218
118 GGAATCTCTGGAAGAAACCTTACGCAACAACT-CAAGGAGAGAAATGAGGAGTAGCAC 176
219 GGCATCTCTGGAAGAAACCTTATGCAACAACAACT-CAAGGAGAGAAATGAGGAGTAGCAC 278
177 GCGCTCATCTGGAAGAAACCTTAAATTTGTGATGCTTCAACAATCTGGAACACTCAA 236
279 GAGCTTCTATCTGGAAGAAACCTTAAATTTGTGATGCTTCAACAACCTGAGGACTCAA 338
237 TGGGAATGCAATTCATATGATCTAGGATGCAACCTCTGGGAGACATGACTGGTGAAGAG 296
339 TGGGAATGCAATTCATATGATCTAGGATGCAACCTCTGGGAGACATGACTGGTGAAGAG 398
297 TGATATCTTTGATGGTCTCTGAGAGTTCCGAGCAATGGCAGAGAAATGCTCATATA 356
399 TGATGCTTTGATGAGTCTCTGAGAGTTCCGAGCAATGGCAGAGAAATATCATATA 458
357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGTGAGCTGGAGAGAGGCTGTGTTA 416
459 AGTCARACCTTAATCGATATTGCTTGTGATCTGTGAGCTGGAGAGAGGCTGTGTTA 518
417 CTGAAGTGAATACCAAGGTTCTTCTGCTGCTTGTGGCTTGTGGCTTGTGGGGGCC 476
519 CTGAAGTGAATACCAAGGTTCTTGTGGTCTTGTGGCTTGTGGCTTGTGGGGGCC 578

QY 477 TGAAGCACAACACTGAAGCTAATAAAGCAGGAAAGCTGGTCTCTGAGTGCACAGAACTTGG 536
Db 579 TGAAGCACAACACTGAAGCTAATAAAGCAGGAAAGCTGGTCTCTGAGTGCACAGAACTTGG 638
QY 537 TAGATTGCTCAACTGAAATATGGAATATGGAAGCTGCAATGGGCTTTCATGACAACTG 596
Db 639 TGAATTGCTCAACTGAAATATGGAATATGGAAGCTGCAATGGGCTTTCATGACAACTG 698
QY 597 CTTTCAGTATATATTGATATCAACAGGCGATGATGATGATGATGATGATGATGATGATGAT 656
Db 699 CTTTCAGTATATATTGATATCAACAGGCGATGATGATGATGATGATGATGATGATGATGAT 758
QY 716 CCATGAATGGAAGTGCAGATATGATCAAAAAGAGAGCTGCCACATGTTCAAAGTATA 716
Db 818 CCATGAATGGAAGTGCAGATATGATCAAAAAGAGAGCTGCCACATGTTCAAAGTATA 818
QY 776 CTGAACCTCTCTTGGCAGTGAAGATGCTTAAAGAGAGCTGGGCCAATAAAGGCCAG 776
Db 878 CTGAACCTCTCTTGGCAGTGAAGATGCTTAAAGAGAGCTGGGCCAATAAAGGCCAG 878
QY 836 TGCTGTGGCTATAGATGCGAGCAGCTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 836
Db 938 TGCTGTGGCTATAGATGCGAGCAGCTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 938
QY 896 ATGAACCATCTGCTACTCAAAATGTAATCATGAGTATGATGATGATGATGATGATGATGATGAT 896
Db 998 ATGAACCATCTGCTACTCAAAATGTAATCATGAGTATGATGATGATGATGATGATGATGATGAT 998
QY 956 TTAATGGGAAGACTACTGCTTGTGAAAAGAGAGCTGGGCCAATAAAGGCCAG 956
Db 1058 TTAATGGGAAGACTACTGCTTGTGAAAAGAGAGCTGGGCCAATAAAGGCCAG 1058
QY 1016 GATATATTGCGATGGCAAGAAACAGTGAATCACTGCGAATGCTGCGAATGCTGCGAATGCTGCGA 1016
Db 1118 GATATATTGCGATGGCAAGAAATGAAGAAATCAATGCTGGGATGCTGCTGCTGCTGCTGCTGCT 1118
QY 1072 ACCGAGAATCTA---GACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1072
Db 1178 ACCGAGAATCTAAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1178
QY 1122 CTCCTTAATTTAATTTTACCTGCTAIA-AIAGIAAAAAIAAAIUTGTCATCA 1122
Db 1229 CTCCTTAATTTAATTTTCTGCTATATCCAGAGAAATATTTGCTCATGA 1229

RESULT 3
ABL61836
ID ABL61836 standard; DNA; 1763 BP.
XX ABL61836;
AC ABL61836;
DT 15-MAY-2002 (first entry)
XX 15-MAY-2002 (first entry)
DE Colon adenocarcinoma related gene sequence SEQ ID NO:173.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
XX gene; ds.
OS Homo sapiens.
XX WO200194629-A2.
PN 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US10838.
PF 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-23313P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.

QY	1073	CTCTTAATTTAAATTTTACCTGCTATA-ATACATAAAATAAAATGTGTGATCA 1122
DB	1179	CTCTTAATTTAAATTTTACCTGCTATAATCCAGACAAATAATTTCTGTCATCA 1229
RESULT 4		
ABQ58199		
ID	AHQ58199	standard; cDNA: 431 BP.
AC	AHQ58199;	
XX	02-AUG-2002	(first entry)
DE	Human colon cancer related nucleotide sequence SEQ ID NO:1894.	
DE	Human; colon cancer; cancer; tissue profiling; forensic; mapping;	
KW	genetic analysis; diagnostic; antisense therapy; gene; ss.	
KW	Homo sapiens.	
OS	Homo sapiens.	
XX	W0200229086-A2.	
PN	11-APR-2002.	
PD	02-OCT-2001; 2001W-(US)30732.	
PF	02-OCT-2000; 2000US-237271P.	
PK	(FARB) BAYER CORP.	
PA	Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;	
PI	Thiagalingam A, Lewis ME;	
PI	WPI; 2002-426115/45.	
XX	New isolated nucleic acid that is differentially expressed in cancer	
PT	tissues useful for determining the presence of colon cancer in a cell	
PT	or tissue type, and in antisense therapy	
XX	Claim 1; Fig 1; 796pp; English.	
PS	ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially	
XX	expressed in cancer tissues. ABB78993 to ABB79004 represent proteins	
CC	encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be	
CC	used in antisense therapy. An antibody immunoreactive with a polypeptide	
CC	encoded by (I) is useful for detecting cancer in a patient sample, and	
CC	for detecting the presence or absence of a polynucleotide encoded by a	
CC	nucleic acid which hybridises to (I) in a cell. A probe/primer derived	
CC	from (I) can be used for determining the presence of a nucleic acid which	
CC	hybridises to (I), and for determining the phenotype of cells in a sample	
CC	of cells from a patient. (I) is useful for determining the presence of	
CC	colon cancer in a cell or tissue type, for determining the presence or	
CC	state of other type of cancer, in antisense therapy, to generate	
CC	macroarrays on a solid surface, to identify a chromosome on which the	
CC	corresponding gene resides, and in tissue profiling, forensics, genetic	
CC	analysis, mapping and diagnostic applications. (I) can be used to raise	
CC	antibodies, and to screen for peptide analogues and antagonists.	
XX	Sequence 431 BP; 122 A; 84 C, 119 G; 106 T; 0 other;	
SQ		
Query Match		
Host	Local Similarity	Score 368.6; DB 24; Length 431;
Matches	392; Conservative	0; Mismatches 39; Indels 0; Gaps 0;
QY	175	ACGGGCTCTCATCTGGGAAAAAATCTTAAATTTGTGATGCTTACAAATCTGGACACTC 234
DB	1	ACGAGCTCTCATCTGGGAAAAAATCTTAAATTTGTGATGCTTACAAATCTGGAGCATTC 60
QY	235	AATGGGAATGCATTCATGATCATACCGATCAACCATCTGGGAGACATGACTGGTGAAGA 294
DB	61	AATGGGATGCACTCATACGATCTGGGATGACCAACCTGGGAGACATGACCTGAAGA 120

bone. Antibodies, or their immunologically active fragments can be used against cathepsin K or nucleic acid encoding cathepsin K as diagnostic markers for breast cancer and its metastases. The antibodies and antisense RNA can also be used against the cathepsin K gene or antagonists for treating breast cancer (metastases).

xx
SQ Sequence 990 BP; 287 A; 210 C; 270 G; 223 T; 0 other;

	Query Match	27.5%	Score 310;	DB 19;	Length 990;
	Best Local Similarity	62 1%;	Pred. No. 5.7e-81;		
	Matches 580;	Conservative 0;	Mismatches 335;	Indels 19;	Gaps 5;
QY	105	CTGGATCATCACTCGAATCTCTGCAACAAAACCTACAGCAAACAAT-CAAGCAACAGAAT	163		
Db	64	CTGGACCCACTGGGAGCTATGGAGAGACCACAGGAAGCATAATACAAACAAGGTG	123		
QY	164	CAGGAAGTAGACAGGCTCTCATCTGGSAAAAAAATCTAAATTTTGTAATGCTTCACAAAT	223		
Db	124	GATGAATCTCTCGGCCGTTTTAATTGGGAAAAAACCTGAAGTATATTTCATCCATAAC	183		
QY	224	CTGGNACACTCAATGGCAATGCTTATATATGATCTAGGCATGACCATCTGGGAGACATG	283		
Db	184	CTTAGGCTCTCTTGTTGGTCCATACATATGAATGCTATGAACACCTGGGAGACATG	243		
QY	284	ACTGCTGAAGAAGTCATATCTTTGATGGGTTCCTCAGAGTTCCC---AGCCAATGCCAG	340		
Db	244	ACCATGGAGAGTGGTTTCAGAGATGACTGGACTCAAGTACCCCTGTCTCATTCGCCG	303		
QY	341	AGAAATGTCAC---TTATAGTCCAACTCTTAATCAGAAATTCGCTGATTTCTGTGGACTGG	397		
Db	304	ACTAATGACACCCCTTATATCCAGAAATGGGAAGGTAGAGCCCCAGCACCTCTCTGCATAT	363		
QY	398	AGAGAAAGGCTGTGTATTATGAAGTCAAAATACACAGTCTTCTGTGGTCTGTGGGCT	457		
Db	364	CGAAAGAAAGATATGTTACTCTGTCTCAAAATAGSGHTCAGSTGAGTCTGTGGGCT	423		
QY	458	TTCAGCGCTGTGGGGGCCCTGGGAAGCACAACTCAAGCTIAAAACAGGAAGCTGTGTCT	517		
Db	424	TTTAGCTCTGTGGTGGCTTGGAGGCCAACTCAAGAGAAACTGGCAAACTCTTAAT	483		
QY	518	CTGAGTGAA'AA'ATTAATTAATTTAACTGAAAAATATGGAAATTAAGAGCTGCAT	577		
Db	484	CTGACTCCCCAGAACC'TAGTGGATTGTGTCTCTCA-----GAATCATGGCTGTGGA	534		
QY	578	GGCGTTTTCATGACAACTGCTTCCAGTATATTATTATTAACACAGGCATTTGATTCAGAA	637		
Db	545	GGGGG'ATATGATCAATGTCTTCCAAATATGTGCAGGAAGACGGGGTATTTGACTCTGA	594		
QY	638	GCTTCTCTATCCCTACAAGGCATCAATGGAAAGTCAGATATGACTTCAAAAAGCGAGCT	697		
Db	595	GATGCTTACCCTATGTGGACAGGAAGAGAGTTGTATGTACAAACCAACAGCGCAANGCA	654		
QY	698	GCCACATGTTCAAAGTATACTGAACCTTCCCTTTGGCAGTGAAGATGCCTTAAAGAAAGCT	757		
Db	655	GCTAAATGCAGCGGTACAGAGATATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA	714		
QY	758	GTGGCCATAAAGACACTGTCTGTGGCTATAGATGGGAGCACCATTTCTTTCTTCCTG	817		
Db	715	GTGGCCCGAGTGGGACCTGTCTGTGGCCATTGATGCAAGGCTCAACCTCCCTCCAGTTT	774		
QY	818	TACGAAGTGGTGCTACTATGAACCATCCG---TACTCAGAATGTGAATCATGCACTA	874		
Db	775	TA'A'A'AAA-RGTGTATTATGAT'AAAG'TGG'AATATG'SATAA'TGTAA'CATGCGGTT	834		
QY	875	TTAGTGTTHGCTATGATTAAC'CTTAATGGGAAGACATACTGGCTGTGAAAAACAGCTGG	934		
Db	835	TTGGCAGTGGGATATGGAATCCAGAGGGCAACCAAGCACTGGATAATTAAAAACAGCTGG	894		
QY	935	GSCCTCAACTTTTGTGA'AAAGATATATTCATGGCAAG'AAATAGTGGAAATCACTGT	994		
Db	895	GAAGAAACTTGGGAAGAA'AAAGGATATATGCTCATGGCTCGAAATTAAGAAACAGCCCTG	954		
QY	995	GGGATTGCTAGTTATCCCTCTTACCACAGAAATCT	1028		

Db	955	GGCATTGCCAAGCTTGGCGACGCTTCCTCCCAAGACTCT	988
RESULT 6			
AAT27040			
ID	AAT27040	standard, DNA: 1482 BP.	
XX	AC		
XX	AAT27040;		
XX			
DT	19-OCT-1996	(first entry)	
DE	Prepro-cathepsin-O2 gene.		
XX			
KW	Human; prepro-cathepsin-O2; cathepsin-O2; spleen-		
KW	spinal disc; adhesion; scar; keloid; cancer; endo-		
KW	pycnodysostosis; bone disorder; therapy; pro-reg-		
KW	cathepsin-O2-inhibitor; diagnosis; antibody, hum-		
XX			
OS	Homo sapiens.		
Key	Location/Qualifiers		
FH	142..1131		
FT	/*tag= a		
FT	/product= prepro-cathepsin-O2		
FT	484..1131		
FT	/*tag= b		
FT	/product= Mature cathepsin-O2		
PN	W09613523-A1.		
XX			
PD	09-MAY-1996.		
XX			
PF	26-OCT-1995; 95W0-US13820		
PR	02-OCT-1995; 95US-0536861.		
PR	27-OCT-1994; 94US-0330121.		
XX			
PPA	(KHEP-) KHEPRI PHARM INC.		
XX			
PI	Broemme D, Okamoto K;		
PI	WPI: 1996-239452/24.		
DR	P-PSDB; AAR95599.		
DR			
PPT	Recombinant human cathepsin O2 protein and relat-		
PPT	ed as a collagenase in vitro and in vivo, e.g.,		
PPT	problems and to dissolve the matrices around tum-		
XX			
PS	Claim 14; Fig 1; 82pp; English.		
XX			
CCC	This sequence encodes human prepro-cathepsin-O2		
CCC	undergoes post-translational processing similar		
CCC	cathepsin-B and -S and papain, with cleavage of		
CCC	sequences to form the mature enzyme. A cDNA pro-		
CCC	from using PCR primers AAT27041-42, and is used		
CCC	spleen cDNA library to obtain a full-length clone		
CCC	has been amplified by PCR using primers AAT27043		
CCC	characterisation of activity. The enzyme has co-		
CCC	in vitro and in vivo, and may be used e.g., in th-		
CCC	problems, pelvic or post-surgical adhesions, sca-		
CCC	endometriosis, pycnodysostosis or aberrant bone		
CCC	pro-region has cathepsin-O2-inhibitor activity.		
CCC	treat cancer and bone disorders, or as a diagnos-		
CCC	Antibodies against the protein may be used diagn-		
XX			
XX	Sequence 1482 BP; 389 A; 350 C; 377 G;366 T; 0		

Query Match 27.5%; Score 310; DB 17; Length 1482;
Best Local Similarity 62.1%; Pred. No. 6.9e-81;
Matches 580; Conservative 0; Mismatches 335; Indels 19

Tue Jul 1 14:40:54 2003

```

xx Key Location/Qualifiers
FH 1 q60
FT /tag= a
FT /product= "Mutant human cathepsin K C139S"
FT misc_feature 1..45
FT /tag= b
FT /note= "Encodes the pre-domain"
FT misc_feature 46..342
FT /tag= c
FT /note= "Encodes the pro-domain"
FT mat_peptide 343..987
FT /tag= d
FT mutation 415..417
FT /tag= e
FT /note= "Changed from TGT in wild-type to TCT in mutant"

W09820024-A1.
14 MAY-1998.
03-NOV-1997: 97WO-CA00824.
04-NOV-1996: 96US-0030411.
(MERI ) MERCK FROSST CANADA INC.
Desmarais S, Friesen R, Zamboni R;
WPI: 1998-348101/30.
P-PSDH: AAW48938.

peptide(s) useful in binding assays for tyrosine phosphatases or
cysteine proteases - contain two or more 4-phosphono(difluoromethyl)
phenylalanine groups to improve binding affinity

Disclosure: Page 41; 59pp; English.

The present sequence represents a mutated human cathepsin K cDNA.
The cDNA encodes a mutant human cathepsin K C139S protease. The
invention provides a method for use in a scintillation proximity
binding assay (SPA) for proteases and phosphatases. The method
involves using mutated proteases and phosphatases whereby the catalytic
cysteine residue of the enzymes are replaced with a serine or alanine
residue to correct the problem of interference in SPA from extraneous
oxidising and alkylating agents. The mutation affects the catalytic
properties of the enzyme but does not affect their binding properties.
The invention claims for new ligands for use in SPA which have
increased binding affinity for a tyrosine phosphatase or cysteine
protease. The ligands contain at least two 4-phosphono(difluoromethyl)
phenylalanine groups which increase binding affinity of the ligand to
its respective enzyme. The assay can be used to determine the ability
of new ligands and compound mixtures to competitively bind with an
enzyme. The method is claimed to allow a better usage of SPA in the
discovery of compounds for the treatment and study of diseases, e.g.
diabetes, cancer and osteoporosis.

Sequence 990 BP: 287 A; 211 C; 269 G; 223 T; 0 other;

Query Match 27.4%; Score 308.4; DH 19; Length 990;
Host local Similarity 62.0%; Pred. No. 1 7e-80;
Matches 579; Conservative 0; Mismatches 336; Indels 19; Gaps 5;

105 CTGATCATCACTGGAAATCTCTGGAGAGAAACCTACAGCAACAAAT-CAAGAGAGAGAAAT 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 CTGACACCCCACTGGAGCTATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 GAGCACTAGCAGGGCTCTCATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 GATGAATCTCTGGAGCTTTAATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 CTGGAACACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGAGAGAGATG 283

```

```

Db 184 CTTGAGGCTTCTCTGGTGTCCATACATATGAACCTGGCIATGAACACCTGGGGACATG 243
QY 284 ACTGGTGAAGAAAGTATATCTTTGATGGTTCCCTGAGAGTTCCG---AGCAATGGCAG 340
Db 244 ACCAGTGAAGAGCTGCTTTCAGAAGATGACTGACCTCAAGTACCCCTGTCTCATCTCCCG 303
QY 341 AGAAATGTCAC---TTATAGGTCAAACTTAATCAGAAATTCCTGATTTCTGTGGACTGG 397
Db 304 AGTAATGACACCCCTTTATATCCAGNATGGGAGGTAGAGCCAGAGACTCTGTGACTAT 363
QY 398 AGAGAGAAGGGCTGTGTTACTGAAGTGAATACCAAGCCTTTCTGTGGTCTGTGGCT 457
Db 364 CGAAACAAAGGATATCTTACTCTGCAAAAATCAGGCTCAGTGGTCTCTGTGGCT 423
QY 458 TTCAGCGCTGTGGGGCCCTGGAAGCACAACTGAAGCTAAAAACAGAAAGCTGTCTCT 517
Db 424 TTTAGCTGTGGTGGTGGCTGGAGGCCAACTCAAGAGAAACCTGCGAACTCTTAAT 483
QY 518 CTGAGTCCACAGAACTTGGTAGATTGCTCAACTGAAATATGGGAATTAAGAGCTGCAAT 577
Db 484 CTGAGTCCCAAGAACCTAGTGGATTGTGTCTGA-----GAATGATGGCTGTGGA 534
QY 578 GCGCGTTTCATGACAACTGCTTCCAGTATATTATGATAACAACGGCATTCATTCAGAA 637
Db 535 GGGGGCTACATGACCAATGCTTCCAAATATGTCACAAGAACCGGGTATTCACTCTGAA 594
QY 638 GCTTCTCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATGACTCAAAAAGCGAGCT 697
Db 595 GATCCCTACCCATATGTGGCAGAGAGAGAGTGTATGTACAAACCAAGCAAGGCA 654
QY 698 GCCACATGTTCAAGTATATCTGAACCTTCCCTTTGGCAGTCAAGTCCCTTAAAAAGAGCT 757
Db 655 GCTAAATGACAGGGTACAGAGAGATCCCGAGGGGATGGAAGAGCCCTGAAGAGGCA 714
QY 758 GTGGCCAATAAGACCTGTCTGTGGCTATAGATCGGAGCCACTATTCTTTCTTCCTG 817
Db 715 GTGGCCCGAGTGGGACCTGTCTGTGGCAATTTGATGCAAGCCCTGACCTCTTCCAGTT 774
QY 818 TACAGAAGTGTGTCTACTATGAACCATCTG---TACTCAGAAATGTGAATCATGGAGTA 874
Db 775 TACAGCAAGGTGTGTTATTTATGTAAGCTGCAATAGCCATATCTGAACCATGCGCT 834
QY 875 TTAGTGGTGGCTATGTTAACTTTAATGGGAAAGCTACTGCTGTGTGAAACACAGCTG 934
Db 835 TTGCGAGTGGATATGGAATCCAGAAGGGAACAAAGCACTGGATAATTAACACAGCTG 894
QY 935 GGCCTCAACTTTGCTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCATCT 994
Db 895 GGAGAAACCTGGGGAACAAAGGATATATCTCTCATGCTCGAATAAGAACACGCTGT 954
QY 995 GGGATTGCTAGTTATCCCTCTTTACCCAGAAATCT 1028
Db 955 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 988

RESULT 12
AAV32616
ID AAV32616 standard; cDNA; 990 BP.
XX AC AAV32616;
XX XX
DT 26-OCT-1998 (first entry)
XX XX
DE Mutant human cathepsin K (C139S) cDNA.
XX XX
KW Mutant human cathepsin K; protease; SPA; tyrosine phosphatase;
KW scintillation proximity binding assay; diabetes; cancer;
KW 4-phosphono(difluoromethyl) phenylalanine group; osteoporosis; ss.
XX OS Homo sapiens.
OS Synthetic.
XX XX

```


100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 08:53:50 ; Search time 1896 seconds

(without alignments)
9618.206 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1126

Sequence: 1 gcaattatagataggagca.....aataaatgtgtcagccat 1126

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_hic:**

9: gb_estl:**

10: gb_est2:**

11: gb_hic:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pin:**

21: em_gss_vit:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715.2	63.5	946	9	AL513990
2	702.4	62.4	892	9	AL548051
3	692.2	61.5	997	9	AL544870
4	682.2	60.6	903	9	AL551671
5	662.6	58.8	1287	11	BC011104
6	572.2	50.8	839	9	AU140074

7	565	2	50	2	848	12	BT539651	602567596
8	562.6	50.0	49.7	1013	14	BQ054160	AGENCOURT	
9	560	49.7	1068	14	BQ710837	14	AGENCOURT	
10	559.8	49.7	1068	14	BM06785	14	AGENCOURT	
11	557.6	49.5	795	12	HC541942	12	AGENCOURT	
12	542.8	48.2	782	12	BG548684	12	AGENCOURT	
13	538	47.8	729	10	AV716038	10	AGENCOURT	
14	535	47.5	729	13	BT766075	13	AGENCOURT	
15	520.6	46.2	1048	10	BE616367	10	AGENCOURT	
16	519.2	46.1	799	10	BE616203	10	AGENCOURT	
17	513.4	45.6	804	9	BQ309582	9	AGENCOURT	
18	513	45.6	804	9	AU138806	9	AGENCOURT	
19	512.8	45.5	642	14	BQ309579	14	AGENCOURT	
20	512	45.5	640	14	BQ309581	14	AGENCOURT	
21	509	45.2	783	13	BM007393	13	AGENCOURT	
22	501.4	44.5	853	10	BE531143	10	AGENCOURT	
23	500	44.4	954	12	BT795515	12	AGENCOURT	
24	485.4	43.1	646	14	BQ327134	14	AGENCOURT	
25	484.4	43.0	729	10	AV683234	10	AGENCOURT	
26	478.8	42.5	937	14	BQ920640	14	AGENCOURT	
27	478.6	42.5	735	12	BT715701	12	AGENCOURT	
28	477.8	42.4	612	12	BF044942	12	AGENCOURT	
29	476.6	42.3	978	9	AV225053	9	AGENCOURT	
30	475.8	42.3	921	14	BQ322641	14	AGENCOURT	
31	463.8	41.2	819	13	BT656302	13	AGENCOURT	
32	461	40.9	668	14	BQ309567	14	AGENCOURT	
33	459.8	40.8	824	13	BT157145	13	AGENCOURT	
34	459.6	40.8	644	14	BM820138	14	AGENCOURT	
35	458	40.7	978	10	BP614890	10	AGENCOURT	
36	457	40.6	866	13	BT694504	13	AGENCOURT	
37	455	40.4	877	12	BF340573	12	AGENCOURT	
38	454.4	40.4	888	14	BQ960433	14	AGENCOURT	
39	453.2	40.2	572	13	BG938409	13	AGENCOURT	
40	451.6	40.1	650	12	BG878075	12	AGENCOURT	
41	449.2	39.9	596	13	BM511045	13	AGENCOURT	
42	445.6	39.6	796	12	BE879042	12	AGENCOURT	
43	445	39.5	965	10	BB611859	10	AGENCOURT	
44	442.2	39.3	598	14	BQ309576	14	AGENCOURT	
45	434.8	38.6	761	12	BG863994	12	AGENCOURT	

ALIGNMENTS

RESULT 1
AL513990
LOCUS AL513990 LTI_NFL006.PL2 Homo sapiens cDNA clone CL0BA0112B04 5
DEFINITION AL513990 946 bp mRNA linear EST 13-FEB-2001
prime, mRNA sequence.
ACCESSION AL513990
VERSION AL513990.1 GI:12777484
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 946)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1. 946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA0112B04"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1. Not 1: 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com 3 others
BASE COUNT 281 a 206 c 243 g 264 t
ORIGIN

Query Match 61.5%; Score 692.2; DR 9; Length 997;
Best Local Similarity 86.4%; Pred. No. 7.5e-191;
Matches 784; Conservative 3; Mismatches 116; Indels 4; Gaps 2;

Db	315	TGCACAGAAATATCACATATAAGTCAAAACCCCTAATTCGGATATTCGCTGATTCCTGGAC	374
Qy	395	TGAGAGAGAGGGCTGTGTACTGAAGTGAATACACAGGGTTCCTGTGGTGTGTGGTGG	454
Db	375	TGGAGAGAGAGGGTGTGTACTGAAGTGAATATCAAGGTCTTGTGGTGTGTGGTGG	434
Qy	455	GCTTTCACGCGTGTGGGGCCCTGGAGACACACATCAAGCTAAACACAGAAAGCTGGT	514
Db	435	GCTTTCAGTGTGTGGGGCCCTGGAGACACACATCAAGCTAAACACAGAAAGCTGGT	494
Qy	515	TCTCTGAGTGCACAGAACTTGGTAGATTCCTCAACTGAAAAATATGGGAATAAAGCTGC	574
Db	495	TCTCTCAGTGCACAACTGGTGGATTCCTCAACTGAAAAATATGGGAATAAAGCTGC	554
Qy	575	AATGGCGTTTCATGACAACTGCTTCCAGTATATTTATGATACACAGCGATTGATTC	634
Db	555	AATGGTGGCTTCATGACAAAGCGCTTCCAGTACATCATTTGATGACAGGCGATCGACTCA	614
Qy	635	GAAGCTTCCTATCTCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAACGGA	694
Db	615	GACGCTTCCTATCTCTACAAAGCCATGGATCAGAAATGTCAATATGACTCAAAATATCGT	674
Qy	695	GCTGCGACATGTTCAAGTATACGAACTTCCTTGGCAGTGAAGATGCTCTTAAAGAA	754
Db	675	GCTGCGACATGTTCAAGTATACGAACTTCCTTATGGCAGAGAGATGCTCTGAAAGAA	734
Qy	755	GCTGTGGCCAAATAAGGACCTGTGTCTGTGGCTATAGATGCGAGCCACTATCTTCTTC	814
Db	735	GCTGTGGCCAAATAAGGACCGCCAGTCTGTGTGGTGTAGATGCGCGCTCATCTCTTCTTC	794
Qy	815	CTGTACAGAGTGGTGTCTACTATGAACCATCTGTACTCAGAAATGTGAATCATGGAGTA	874
Db	795	CTGTACAGAGTGGTGTCTACTATGAACCATCTGTACTCAGAAATGTGAATCATGGTGA	854
Qy	875	TTAGTGGTGGCTATGATGATACCTTAATGGGAAGACTA	912
Db	855	CTTGTGGTGGCTATGATGATCTTAATGGGMAAGATA	892
RESULT 3			
AL544870			
LOCUS			
DEFINITION AL544870 LTI_NFL006_PL12 Homo sapiens cDNA clone CS0D1012YD10 5			
prime, mRNA sequence.			
ACCESSION AL544870			
VERSION AL544870.1 GI:12877351			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE			
AUTHORS Li W.B., Gruber C., Jessee J., and Polayes D.			
TITLE Full-length cDNA libraries and normalization			
JOURNAL Unpublished (2001)			
COMMENT Contact: Genoscope			
Genoscope - Centre National de Sequencage			
Bp 191 91006 EVRY cedex - France			
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.			
FEATURES			
source			
1..997			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="CS0D1012YD10"			
/clone_lib="LTI_NFL006_PL12"			
/tissue_type="placenta"			
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA			
was primed with a NotI-oligo(dT) primer. Five prime end			
enriched, double-stranded cDNA was digested with Not I and			
cloned into the Not I and Eco RV sites of the pCMVSPORT 6			
vector. Library was normalized. Library was constructed by			
Life Technologies. Contact : Feng Liang Life Technologies,			
a division of Invitrogen 9800 Medical Center Drive			

```

RESULT 4
LOCUS AL551671
DEFINITION AL551671 L1_NFL006_PL2 Homo sapiens cDNA clone CS0D1062YC04 5
prime, mRNA sequence.
ACCESSION AL551671
VERSION AL551671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Fekaryota, Metazoa; Chordata; Craniata, Vertebrata, Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 903)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Pollayes, D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..903
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1062YC04"
/issue_lib="LTI_NFL006_PL2"
/issue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive,
Rockville, Maryland 20850, USA Fax: (1) 301 610 8371
Email: fulllength@invitrogen.com"
BASE COUNT 258 a 192 c 226 g 226 t 1 others
ORIGIN
Query Match 60.6%; Score 682.2; DH 9; Length 903;
Best local Similarity 87.1%; Pred. No. 6e-188;
Matches 771; Conservative 1; Mismatches 109; Indels 4; Gaps 2:
QY 1 GCATTATTAGTATGGGAGACCTGACGTTTCTTTTC---ATGAATGGGTAGTTGGGTGC 57
DB 19 GGACTCTTACCTGGGAGCACTGCTGCTTATCACAATGAACGCGCTGTTGTGTC 78
QY 58 TTCGGTTGTGCTCTATGCTAGTGGCACAAGTACATAAGATCCACTCTGGATCATCACT 117
DB 79 TCTTGGTGTGCTCTGCTAGTGGCAGAGTGCATAAAGATCCTACCTGGATCACCACCT 138
QY 118 GGAATCTCTGGAGAAACCTACAGCAACAAT-CAAGCAGAGCAATGAGGAAGTAGCAC 176
DB 139 GGCATCTCTGGAGAAACCTATGCAACAATATAGCAACAATATAGAGACAGTAC 198
QY 177 GCGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACATCTCAAGCACTCAA 236
DB 199 GACGCTCATCTCGGAAAAAATCTAAATTTGTGATGCTTCACAACTCGGAGCATCAA 258
QY 237 TGGGAATGATTTCATATGATCTAGGCTAGGCAATCCATCTGGGAGACATGACTGGTAGAG 296
DB 259 TGGGAATGACTCTATGATCTGGGCAATGACCACTGGGAGACATGACTGGTAGAGAG 318
QY 297 TCATATCTTTGATGGTTCCCTGAGAGTTCCAGCAATGGCAGCAATGGCAGCAATGTCACCTATA 356
DB 319 TGATGCTCTTTGATGAGTTCCCTGAGAGTTCCAGCAATGGCAGCAATGTCACATATA 378
QY 357 GGTCAAACTCTAATCAGAAATTTGCTGATCTGCTGAGTGGAGAGAGAGGCTGTGTTA 416
DB 379 ACTCAACGCTAATTCGATATTCCTGATTCCTGCTGAGTGGAGAGAGAGGCTGTGTTA 438

```

417 CTGAAGTCAAAATACAGGGTCTCTTGGTCTTCTTGGCTTTTACGGCTTCTACGGCTTCTGGGGCCC 476

439 CTGAAGTCAAAATATCAAGGTCTCTTGGTCTTCTTGGCTTTTACGGCTTCTGGGGCCC 498

477 TGAAGCACAACTCAAGCTAAAAACAGGAGCTGGTCTCTCTCAGCTGCATCAAAATTTGG 536

499 TGAAGCACACGCTCAAGCTCAAAACAGGAGCTGGTCTCTCTCAGCTGCATCAAAATTTGG 558

537 TAGATTGCTTAAATGAATAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 596

559 TGGATTGCTCAACTCAAAATATATGAACAAAGAGCTGCAATGGTGGCTTCAACCAACGG 618

597 GTTTCCTCACTATATTATTGATAACACGCGATTCGATTCAGAGCTTCCTATCTATCAAAAG 656

619 GTTTCCTCACTATATTATTGATAACACGCGATTCGATTCAGAGCTTCCTATCTATCAAAAG 678

657 CCATGAATGGAAGTGCAGATATGACATCAAAAAGGAGCTGGCAGATCTTCAAACTATA 716

679 CCATGGATCAGAAATGTCAATATCATTCAAAATATCTGCTGCGACATGTTCAAACTACA 748

717 CTGAATCTCTCTTGGAGTCAAGATGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 776

739 CTGAATCTCTCTTGGAGTCAAGATGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 798

777 TGTCTGGCTATAGATCGGAGCGCATCTATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 846

799 IGCTGTTGGTGTAGTGGGCTATCT 858

837 ATGAACATCTCTTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 881

859 ATGAACATCTCTTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 903

LOCUS BC011104 1287 bp mRNA linear HTG 07-AUG-2002

DEFINITION Mus musculus, cathepsin S, clone IMAGE:3491724, mRNA

ACCESSION BC011104

VERSION BC011104 1 GI:15029768

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata, Vertebrata, Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1287)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (25-JUL-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 23 Row: c Column: 24

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein

this clone has the following problem: frame shifted.	
FEATURES	Location/Qualifiers
source	1. 1287
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/map="C28H11"
	/clone="IMAGE:3491724"
	/tissue_type="mammary tumor metastasized to lung. Tumor
	arose spontaneously from a senescent normal mammary
	(clonal) outgrowth infected with the virus MMTV."
	/clone_lib="NCI_CGAP_Lu29"
	/lab_host="DH10B"
	/note="vector: pCMV-SPORT6"
BASE COUNT	372 a 269 c 315 g 331 t
ORIGIN	
Query Match	58.8%; Score 662.6; DB 11; Length 1287;
Best Local Similarity	76.6%; Pred. No. 3.6e-182;
Matches	866; Conservative 0; Mismatches 254; Indels 11; Gaps 4;
QY	2 CATTATTAGTATGGAGCACCTGCAGGTCTTTTCATGAATGGCTAGTTGGGCTGTCTCC 61
DB	22 CAGGATGACCATGAGGCTCTCGCCAGGCTGCCATAAGATGGCTGTTTGGATGCCCT 81
QY	62 GTTGTGCTCTATGCGTGGCAGTACATAAAGATCCCACTTGCATCATCACTGGAA 121
DB	82 GGTGTGCTCTGTGCAATGGAGCAACTGCAGAGAGACCTACCTGGACTACCACTGGGA 141
QY	122 TCTCTGGAGAAACCTACAGCAACAT-CAAGGAGAGAAATGAGGAAGTAGCAGCGG 180
DB	142 TCTCTGGAGAAACCCATGACAAGAAATACAGGATGAAGATGAAGAAGAAAGTACGCGG 201
QY	181 TCTCATCTGGGAAAAATCTAAATTTGTGATGCTTCACAACTCGGACACTCAATGGG 240
DB	202 TCTCATCTGGGAAAAAGATCTGAAGTTTTATAATGATTCACAACTCGAATTTCTATGG 261
QY	241 AATGCATTATGATGCTAGGATCAACCATCTGGGAGACATGACTGGTGAAGAAGTAT 300
DB	262 AATGCATCTACCAAGTGGGATGACGATATGGAGACATGACCAATGAAGAATCTT 321
QY	301 ATCTTTGATGGTTCCTGAGATGCCAGCCAAATGGCAGAGAAATGCTACTATAGGTC 360
DB	322 GTGTCGATGGGTCTCTGAGAAATCCCGCAGACTCTCCGAGACTGTCATCTTCAGTCT 381
QY	361 AAATCTAATCAGAAATTCCTGATCTCTGGAATGGAGAGAGAGAGAGGCTGTCTTACTGA 420
DB	382 ATACTCTAATCGACATTCCTGACACTGTGGACTGGAGAGAGAGAGAGGCTGCGTCACTGA 441
QY	421 AGTCAATACCAAGGTTCTTGTGGTCTCTGTTGGGCTTTCAGCGCTGTGGGGCCCTGGA 480
DB	442 GGTGACATAAAGGGTCTTGTGGTGGCTTTCAGTGTGCTGTGGGGCCCTTGA 501
QY	481 AGCAACTGAGCTAAACAGCAAGTGTGTCTCTGAGTGCACAGAACTTGGTAGA 540
DB	502 AGGCAGCTGAACCTCAAAACGGGAGCTGATATCCCTCAGTGTCTCAGAACCTGTGTGA 561
QY	541 TTGCTCAACT---GAAAAATATGGGATTAAGAGCTGCAATGCGGTTTTCATGACAACTGC 597
DB	562 CTGCTCAATGAAGAAAGTATGGGAATAAAGCTGTGGAGGCGGTACATGACCGAAGC 621
QY	598 TTTCAGTATATTATTCATAAACAGGCAATTTGATTCAGAAAGCTTCCTATCCCTTCAAAAGC 657
DB	622 TTTCAGTATATTATTCATAAACAGGCAATTTGATTCAGAAAGCTTCCTATCCCTTCAAAAGC 681
QY	658 CATGAATGGAAGTGCAGATATGACTCAAAAAGCGAGCTGCACATGTTCCAAAGTATAC 717
DB	682 CATGATCAAAAGTGTCTACTATACTCAAAAATTCGAGCTGCTACGTGTTCAAGGTACAT 741
QY	718 TGAATTCCTTTTGGCACTGAAGATGCTTAAAGAAAGCTGTGGCCAAATAAAGGACCTGT 777
DB	742 TCAGTCTCCCTTTTGGTACGAGGATGCCCTGAAAGAAAGCAGTGGCCACTTAAAGGGCCTGT 801
QY	778 GTCCTGTGCTATAGATCGGAGCCACTATCTCTTTCTCTGTACAGAAAGTGGTGTCTACTA 837


```

Db      347  TGAATGCTTTGATGAGTTGCTGAGAGTTCCCAAGCAGTGGCAGAGAAATATACATATA 406
Qy      357  GGTCAAACTCTAATCAGAAATTCGCTGATTCCTGGACTGGACAGAGAAAGGGGTGTGTTA 416
Db      407  AGTCAAACTCTAATCAGAAATTCGCTGATTCCTGGACTGGACAGAGAAAGGGGTGTGTTA 466
Qy      417  CTGAAGTGAATAACAGAGGTTCTGTGTGTTGTTGGCTTTTCAGGCTGTGGGGGCC 476
Db      467  CTGAAGTGAATAATCAAGGTTCTGTGTGTTGTTGGCTTTTCAGGCTGTGGGGGCC 526
Qy      477  TGGAGACACAACTCAAGCTCAAAACAGGAAGTGTGTCTCTGAGTGCACAGAACTTGG 536
Db      527  TGGAGACACAGCTGAAGCTCAAAACAGGAAGTGTGTCTCTGAGTGCACAGAACTTGG 586
Qy      537  TAGATTCCTCAACTGAAATATATGGGAATAAAGGCTGCAATGAGCGGTTTCATGACAACTG 596
Db      587  TGGATTCCTCAACTGAAATATATGGGAATAAAGGCTGCAATGAGCGGTTTCATGACAACTG 646
Qy      597  CTTTCCAGTATATTATTGATCAACAGCGCATTTGATTCAGAGGCTTCCTATCCCTACAAAG 656
Db      647  CTTTCCAGTATATTATTGATCAACAGCGCATTCGACTCAGAGCTTCCTATCCCTACAAAG 705
Qy      657  CATTCAATCAATGTCAGATATGACTCAAAAGAGGAGCTGGCCACATGTTCAAAGTATA 716
Db      706  CCATGATCAGAAATGTCATGACTCCAAATATGCTGTGCTGACACATGTTCAAAGTATA 765
Qy      717  CTGAATTCCTTTGGCAGTGA--GATGCCCTTAAAGAGAGCTGTGCCCAATAAAGGACC 774
Db      766  CTGAATTCCTTTATGGNAGGAAGATGCTCTGAAAGAAAGCTTGGGCGCAATAAAGGCC 825
Qy      775  TGTGTC 780
Db      826  AGGCC 831

```

RESULT 11

```

BG541942
LOCUS      602569445Fl NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4694098 5',
DEFINITION mRNA sequence.

```

```

ACCESSION BG541942.1 GI:13534175
VERSION   EST.
KEYWORDS  human.
SOURCE    Homo sapiens

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

NIH-MGC help://mrg.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov

```

```

Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```

Plate: LLCM1517 row: i column: 11
High quality sequence stop: 790.

```

```

Location/Qualifiers
1..795
/oranism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4694098"
/clone_lib="NIH_MGC_77"
/lab_host="PH10B (T1 phage-resistant)"

```

```

/note="Organ: lung; Vector: pDNR-LJH (Clontech); Site_1:
5'fl (ggccgctcgcc); Site_2: 5'fl (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor

```

FEATURES

```

source
1..795
Query Match 49.5%; Score 557.6; DH 12; Length 795;
Best Local Similarity 86.3%; Pred No 1.4e-151;
Matches 675; Conservative 0; Mismatches 99; Indels 8; Gaps 5;

```

```

Qy      9  AGTATGGGAGCACTGCAGGTTCTTC---ATGAATGGCTAGTTGGGCTGCTCGGTG 65
Db      3  ACTGTGGGAGCACTGCTGTTCTATCACAATGAACGGCTGTTGTTGCTCTGTTG 62
Qy      66  TGTCTCTATGCTAGTGGCACAAGTACATAAAGATCCACTCTGGATCATCACTGGAATCTC 125
Db      63  TGTCTCTCTGCACTGGCAGATTGCATAAAGATCCCTACCTGGATCACCACCTGGCATCTC 122
Qy      126  TGAAGAAAAACCTTACAGCAACAAT--CAAGGAGAGCAATGAGCAACTAGCAGCGCTCTC 184
Db      123  TGAAGAAAAACCTTATGGCAACAATACAGGAAAAAGATGAGAGAG--ATACAAATCTC 182
Qy      185  ATCTGGGAAAAAATCTAAAATTTCTCATCTCTCACAATCTCGAACACTCAATGGGAATG 244
Db      183  ATCTGGGAAAAAATCTAAAATTTTGTGATGCTTCACAACCTGGAGCATTCATGGCAATG 242
Qy      245  CATTATATGATCTAGGATGAACCATCTGGGAGACATGACTGGTGAAGAGTATCTCT 304
Db      243  CACTATACCATCTGGGCATGAACCACTGGGAGACATGACCACTGAAGAGTATGATCT 302
Qy      305  TTGATGGTTTCCCTGAGAGTTCCAGCAATGGCAGAGAAATCTCACTTATAGTCAAAAC 364
Db      303  TTGATGAGTTCCCTGAGAGTTCCAGCACTGGCAGAGAAATATCATATATAGTCAAAAC 362
Qy      365  TCTAATCAGAAATTTGCTGATCTGGAGTGGAGAGAGAGGCTGTGTTACTGAAGTC 424
Db      363  CCTAATCGGATATTGCTGATCTGGAGTGGAGAGAGAGGCTGTGTTACTGAAGTC 422
Qy      425  AATACACAGGTTCTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 484
Db      423  AATATCAAGTTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 482
Qy      485  CAACCTGAAGCTAAAAACAGAGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 544
Db      483  CAGCTGAAGCTGAAACAGGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 542
Qy      545  TCAACTGAAAAAATATGGGAA--TAAAGGCTGCAATGGGCGTTCATGAGCAACCTCTTCCA 603
Db      543  TCAACTGAAAAAATATGGGAAAGGCTGCAATGGTGGCTTCATGAGCAAGGCTTTTCCA 602
Qy      604  GTATATTATTGATAAACAGCGCATTTG--ATTGAGAGGTTTCTTATCTCTTAAAGCTATGA 662
Db      603  GTACATCATTCATAACAGGCGCATTCGAACTCAGAGCTTCTCTATCTCTCTCTCTCTCTCT 662
Qy      663  ATGAAACTCCAGATATACATCAAAAAAGAGAGTGGCTGCTCATGTCTTCTCTCTCTCTCTCT 722
Db      663  ATCAGAAATGTCAATATGACTCAAAATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 722
Qy      723  TTCCCTTTGGCAG--TGAAGATGCTTTAAAGAGAGCTGTGGCCCAATAAAGCAGCTGTCTC 780
Db      723  TTCCCTTTGGCAGAGAGAGAGATGCTTCAAGAGAGCTGTGGCCCAATAAAGCAGCTGTCTC 782
Qy      781  TG 782
Db      783  TG 784

```

```

RESULT 12
BG548684

```

sequence: 5'-CACGGCCATTATGCTC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGAGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library.

BASE COUNT 236 a 165 c 204 g 190 t
ORIGIN

1	AAATATCAAGAGTCTTGTGGTGGCTTTCAGTGGCTGTGGGGCCCTGGAAGCA	60
485	CAACTCAAGCTAAAAACAGGAAGCTGGTCTCTAGTCCACAGAACTTGGTAGATTGC	544
61	CAGCTGAAGCTGAAACAGGAAGCTGGTCTCTAGTGGCCAGAACCTGGTGGATTGC	120
545	TCAATGAAAAATATGGGAATAAAGCTGCAATGGGGGTTTCATGACAACTGTTCCAG	604
121	TCAACTGAAAAATATCGAAACAAAGGCTCAATGGTGGCTTCATGACAAAGGCTTTCAG	180
605	TATATTATTGATAAACAGGCATTGATTCAGAAGCTTCTATCCCTACAAAGCCATGAAT	664
181	TACATCATTGATACCAAGGTCATCGACTCAGAGCTTCTCTATCCCTACAAAGCCATGGAT	240
665	GGAAGTCGAGATATGACTCAAAAAAGCAGCTGGCCACATGTTCAAAGTATACTGAACCTT	724
241	CAGAAATGTCATATGATCAATAATATGCTGTGCCACATGTTCAAAGTACACTGAACCT	300
725	CCCTTTGGCAGTGAAGATGCTTTAAAGAAAGCTGTGGCCCAATAAAGAGCTGTGTCTGTG	784
301	CCTTATGGCAGACAAGATGTCCTGAAAGAAGCTGTGGCCCAATAAAGGCCAGTGTCTGTT	360
785	GCTATAGATGCGAGCCACTATTCTTTCTCTCTACAGAGTGGTGTCTACTATGAACCA	844
361	GGTGTAGATGCGGCTATCCCTCTCTCTACAGAGTGGTGTCTACTATGAACCA	420
845	TCTGTACTCAGAAATGTGAATCATGGAGTATTAGTGGTGGCTATGCTAACCTTAATGGG	904
421	TCTGTACTCAGAAATGTGAATCATGGTGTACTTGTGGTGGCTATGCTTAATGGG	480
905	AAAGACTACTGGCTTTGTGAAAAACAGCTGGGGGCTCNACTTTGGTGACCAGGATATATT	964
481	AAAGAAATACTGGCTTTGTGAAAAACAGCTGGGGGCAACACTTTGGTGAAGAGGATATATT	540
965	CGGATGCAAGAAGAACAGTGGAAATCACTGTGGGATTCGTAGTTATCCCTCTTACCCAGAA	1024
541	CGGATGCAAGAATAAAGGAAATCATTTGTGGGATTCGTAGTTTCCCTCTTACCCAGAA	600
1025	ATCTA---GACCTCTCATTTTATACAAAGTCCAAAAAATGAAACACTTTCCTTAATT	1081
601	ATCTAGAGGATCTCTCTTTTATACAAATCAGAATAATGAAGCACTTTCCTTAACCT	660
1082	TAATTTTACCTGCTATAATAGTAAAAATAAATGTGTCATGA	1122
661	AATTNTTCCCTGCTCTATCCAGAGAAATAAATTTGTGTCATGA	701

RESULT 14
 BI766075
 LOCUS
 DEFINITION 758 bp mRNA linear EST 25-SEP-2001
 603053051P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202444 5',
 BI766075
 mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 758)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: rgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

```

Plate: L1AM1507 row: f column: 13
High quality sequence stop: 752.
Location/Qualifiers
1. .758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202444"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: Not1; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens; library is oligo-dt
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026 Note:
this is a NIH_MGC Library."
236 a 161 c 192 q 179 t
BASE COUNT

```

Query Match	47.5%	Score 535.2	DB 13	Length 758
Best Local Similarity	85.6%	Pred. No. 5e-145		
Matches 630	Conservative 0	Mismatches 103	Indels 3	Gaps 3
QY	36	ATGAATATGGCTAGTTGGCGTCTCTCCCGTTGTGCTCCCTATGCAGTGGGCACAGTACATAAA	95	
Db	21	ATGAACGGCTGTTGTG-TGCCTTGGTGTGCTCCCTGCAGTGGCACAGTTGCATAAA	79	
QY	96	GATCCCACTCTGGATATCACTGAATATCTCTGGGAAGAAACCTTACAGCAACAT-CAG	154	
Db	80	GATCCTACCCCTGGATCACTATGGCATCTCTGGGAAGAAACCTATGGCAACAATACAG	139	
QY	155	GAACAGATAGGAAGTAGGACCGCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTGATG	214	
Db	140	GAAGAATGAAGAAGCAGTACAGCT-TTATCTGGGAAGAATCT-AAAGTTTGTGATG	198	
QY	215	CTTCACAACTCTGGAACTCAATATGGGAATGCATTCATATGATCTAGGCATCAACCATCTG	274	
Db	199	CTTCACAACTCTGGAACTCAATATGGGAATGCATTCATATGATCTAGGCATCAACCATCTG	258	
QY	275	GGACATGACTGGTGAAGAAGTGTATCTTTTGTATGGGTTCCCTCGAGATTTCCAGCAAA	334	
Db	259	GGACATGACTGGTGAAGAAGTGTATCTTTTGTATGGGTTCCCTCGAGATTTCCAGCAAA	318	
QY	335	TGGCAGAGAAATGTCACTTATATAGTCAAACTCTAATCAGAAATTCCTTATCTCTGCAC	394	
Db	319	TGGCAGAGAAATGTCACTTATATAGTCAAACTCTAATCAGAAATTCCTTATCTCTGCAC	378	
QY	395	TGGCAGAGAAAGGGTGTGTACTGAAGTGAATACCAAGGTTCTTTGTGGTCTGTTTGG	454	
Db	379	TGGCAGAGAAAGGGTGTGTACTGAAGTGAATACCAAGGTTCTTTGTGGTCTGTTTGG	438	
QY	455	GCTTTCAGCGTGTGGGGGCGCTGGAAGCAAACTGAAGCTTAAACAGGAAGGTTGGT	514	
Db	439	GCTTTCAGCGTGTGGGGGCGCTGGAAGCAAACTGAAGCTTAAACAGGAAGGTTGGT	498	
QY	515	TCCTCAGTGGCACAACCTTGATACATTCCTCAACTGAAAATATGGGAATTAAGCGCTCC	574	
Db	499	TCCTCAGTGGCACAACCTTGATACATTCCTCAACTGAAAATATGGGAATTAAGCGCTCC	558	
QY	575	AATGGCGTTTCATGACAACTGCTTTCAGTATATTTATGATAACACGGCATTCATTTCA	634	
Db	559	AATGGCGTTTCATGACAACTGCTTTCAGTATATTTATGATAACACGGCATTCATTTCA	618	
QY	635	GAAGT-TTCTATCTACTAAATATATGAATATCAAAATGTAATATCAT-TCAAAAAGTGA	694	
Db	619	GACGCTTCTATCCCTACAAAGCCATGATCAGAAATGTCAATATCAGTCAAAATATCGT	678	
QY	695	GCTGCCATGTTCAAAGTATACTGAACCTTCCTTTGGCAGTCAAGATGCCCTTTAAACAA	754	
Db	679	GCTGCCATGTTCAAAGTATACTGAACCTTCCTTTGGCAGTCAAGATGCCCTTTAAACAA	738	

Tue Jul 1 14:40:54 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:15:25 ; Search time 70 Seconds
(without alignments)
4933.115 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1126
Sequence: 1 gcattattagtgaggagca.....aataatgtgtcatgacat 1126

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapexl 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5R_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6R_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCPMIS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	845.4	75.1	1643	US-09-701-685-1	Sequence 1, Appli
2	310	27.5	1482	US-08-330-121B-1	Sequence 1, Appli
3	310	27.5	1482	PCT-US95-13820-1	Sequence 1, Appli
4	310	27.5	1619	US-08-208-007A-1	Sequence 1, Appli
5	310	27.5	1619	US-08-915-095A-1	Sequence 1, Appli
6	310	27.5	1619	US-08-798-096-1	Sequence 1, Appli
7	310	27.5	1619	US-08-798-095A-1	Sequence 1, Appli
8	310	27.5	1619	PCT-US94-04781-1	Sequence 1, Appli
9	310	27.5	1669	US-08-964-308-3	Sequence 3, Appli
10	310	27.5	1669	US-08-964-313-3	Sequence 3, Appli
11	310	27.5	1669	US-09-069-138-3	Sequence 3, Appli
12	308.4	27.4	990	US-08-964-308-12	Sequence 12, Appli
13	308.4	27.4	990	US-08-964-313-12	Sequence 12, Appli
14	308.4	27.4	990	US-09-069-138-12	Sequence 13, Appli
15	306.8	27.2	990	US-08-964-308-13	Sequence 13, Appli
16	306.8	27.2	990	US-08-964-313-13	Sequence 13, Appli
17	306.8	27.2	990	US-09-069-138-13	Sequence 35, Appli
18	306.8	27.2	1614	US-08-684-932A-35	Sequence 1, Appli
19	300.2	26.7	990	US-08-806-959-1	Sequence 2, Appli
20	236	21.0	1366	US-08-883-526-2	Sequence 1, Appli
21	162.2	14.4	651	US-08-546-712-1	Sequence 1, Appli
22	162.2	14.4	651	US-08-751-105-1	Sequence 1, Appli
23	137	12.2	1390	US-08-821-994-61	Sequence 61, Appli
24	137	12.2	1441	US-08-821-994-63	Sequence 63, Appli
25	133.8	11.9	1434	US-08-821-994-62	Sequence 62, Appli
26	132.2	11.7	1102	US-08-821-994-86	Sequence 86, Appli
27	130.2	11.6	1661	US-08-821-994-82	Sequence 82, Appli

ALIGNMENTS

RESULT 1

US-09-701-685-1

; Sequence 1, Application US/09701685

; Patent No. 6387629

; GENERAL INFORMATION:

; APPLICANT: Schneider, Patrick

; APPLICANT: Yamamoto, Karen K.

; APPLICANT: French, Cynthia K.

; APPLICANT: Reprogen, Inc.

; TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of

; TITLE OF INVENTION: Endometriosis

; FILE REFERENCE: 018002-001310US

; CURRENT APPLICATION NUMBER: US/09/701.685

; CURRENT FILING DATE: 2001-03-26

; PRIOR APPLICATION NUMBER: WO PCT/US99/12335

; PRIOR FILING DATE: 1999-06-03

; PRIOR APPLICATION NUMBER: US 60/088,017

; PRIOR FILING DATE: 1998-06-04

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1643

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (70)..(1065)

; OTHER INFORMATION: human cathepsin S

US-09-701-685-1

Query Match 75.1%; Score 845.4; DB 4; Length 1643;
Best Local Similarity 87.2%; Pred. No. 9.5e-251;
Matches 986; Conservative 0; Mismatches 136; Indels 9; Gaps 5;

QY	1	GCATTATTAGTGGAGCACCTGCAGGTTCTTTC-----ATGAATGCTTATTTGGGTTGC	57
DB	32	GGACTCTTACTTAGGAGCAACTGCTGTTCTATCAATGAACGGCTGGTTGTGTGC	91
QY	58	TTCCGTTGTCTCTATGCAGTGGCACAACATCAATGATCCACTCTGGATCATCAT	117
DB	92	TCTTGGTGTCTCTCTCTGTCAGTGGCAGTGGTGTATTAAGATCTTACCCTGGATCCACT	151
QY	118	GGAATCTCTGGAAGAAACCTACAGCAACAAT--CAAGGAAGAGAAATGAGGAGTAGCAC	176
DB	152	GGCATCTCTGGAAGAAACCTATGCGCAACAATACAGGAAGAAATGAAGAGCAGTAC	211
QY	177	GCGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCAATCTGGACACTCAA	236
DB	212	GACGCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCAACACCTGGACACTCAA	271

Sequence 1, Appli
Sequence 1, Appli
Sequence 48, Appli
Sequence 102, Appli
Sequence 16, Appli
Sequence 31, Appli
Sequence 16, Appli
Sequence 31, Appli
Sequence 16, Appli
Sequence 30, Appli
Sequence 14, Appli
Sequence 30, Appli
Sequence 14, Appli
Sequence 30, Appli
Sequence 14, Appli
Sequence 29, Appli
Sequence 64, Appli
Sequence 60, Appli

28 125.2 11.1 1056 1 US-09-500-651-1
29 125.2 11.1 1056 1 US-08-813-591-1
30 124.6 11.1 1600 4 US-03-325-932A-98
31 123.8 11.0 2167 4 US-09-325-932A-102
32 122.6 10.9 1203 4 US-09-005-298-16
33 122.6 10.9 1203 4 US-09-005-298-31
34 122.6 10.9 1203 4 US-08-768-619-16
35 122.6 10.9 1203 4 US-08-768-619-31
36 122.6 10.9 1203 5 PCT-US96-09848-16
37 122.6 10.9 1203 5 PCT-US96-09848-30
38 122.6 10.9 1306 4 US-09-005-298-14
39 122.6 10.9 1306 4 US-09-005-298-30
40 122.6 10.9 1306 4 US-08-768-619-14
41 122.6 10.9 1306 5 PCT-US96-09848-14
42 122.6 10.9 1306 5 PCT-US96-09848-29
43 122.6 10.6 1474 4 US-08-821-994-64
44 119.8 10.6 1553 4 US-08-821-994-60
45 119.8 10.6 1553 4 US-08-821-994-60


```
QY 638 GATTCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATGACTCAAAAAAGCGAGCT 697
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 GATCCCTACCATATGTGGACACAGGAAGAGAGTGTATGTACAAACCCACAGCAAGGCA 795
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 GGCACATGTTTCAAGATATGACTGAACCTTCCCTTTGGCAGTGAAGATCGCTTAAAGAGCT 757
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 GCTAATATGAGAGGTACAGAGATCCCGGAGGGAATGAGAAAGCCCTGGAAGAGGCA 855
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 GTGGCCAAATAAAGACCTGTGTGTGCTTATACATGCGGAGCCACTATTCTTTCTTCCTG 817
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 GTGGCCGAGTGGAGCTGTCTGTGTGGCATTGATGCAAGCTGACCTCTCTCCAGTTT 915
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 818 TACAGAAGTGTGTCTACTATGAACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 916 TACASCAAGGTGTGTTATTAATGAAGCTGCAATAGCGATAATCTGAAACCATGGGTT 975
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 875 TTACTGTGTGCTATGTAACCTTAATGGGAAGACTTACTGGCTTGTGAAAAACAGCTGG 934
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 976 TTGCAGTGGATATGAAATCCAGAGGGAACAGCACTGGATTAATTAACACAGCTGG 1035
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 935 GGCCTCAACTTGTGTGACCAAGGATATATCGGATGCAAGCAACAGCTGGAATCACTGT 994
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1036 GGAGAAACTGGGAAACAAAGGATATATCTCATGGCTGGAATTAAGAAACAGCGCTGT 1095
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1096 GGCAITGCCAACCTGGCCAGCTTCCCAAGATGT 1129
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
PCT-US95-13820-1
; Sequence 1, Application PC/TUS9513820
; GENERAL INFORMATION: Khepri Pharmaceuticals, Inc.
; TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13820
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US UNKNOWN
; FILING DATE: 02-OCT-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/330,121
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: FP-60261-1-PC/DJR/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 142..1128
; PCT-US95-13820-1

Query Match 27.5%; Score 310; DB 5; Length 1482;
Best Local Similarity 62.1%; Pred No 1e-85;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5.

QY 105 CTGGATCATCTACTGGAATCTCTGGGAAGAAACCTCAGCAACAAT-CAAGCAAGAGAAAT 163
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 205 CTGGACACCCACTGGGAGCTATGGAAGAAGAACCCACAGGAACAATATAA-AAAGATG 264
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 GAGGAAGTAGCAGCGCTCTCATCTGGGAAAAAAATCTAAAAATTTGTGATGCTTCACAAAT 223
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265 GATGAATCTCTCGGGGTTTAAATTTGGGAAAAAAACCTCAGATATATTTCCATCCATAAC 324
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 224 CTGGAACACTCAATGGGAATGCATATATGATCTAGGATATGAATATTTGAAATATTTGAAATATG 283
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 325 CTGTAGGCTTCTCTTGTGCTCCATACATATGAACCTGGCTATGAACCCACCTGGGGACATG 384
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 284 ACTGCTCAACAAGTGTATCTTTGAAGGGTTCCTTCAGAGATTCCTTCTCTCTCTCTCTCTCTCT 340
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 385 ACCAGTGAAGAGGTGTTTCAAGAGATGATGAGCTCAAAAGTACCTCTCTCTCTCTCTCTCTCTCT 444
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 341 AGAAATGTCTAC---TTATAGCTCAAACTCTTAATCAGAAATTTGCTCTGATTTCTGTGGACTGG 397
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 445 AGTAATGACACCTTTATATCCAGAAATGGGAAGGTAGAGCCCCACACTGTGTGACTAT 504
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 AGAGAGAAGGGCTGTGTACTGAAGTGAATATCCAGGGTCTTGTGTGGTCTTGTGGCT 457
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 505 CGAAAGAAAGCATATGTTACTCTGTCAAAATATCAGGGTCAAGTGTGCTTCTGTGGCT 564
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 458 TTCAGGCTGTGGGGCCCTGGGAAGCACAACTCAAGCTAAAAACAGAAAGCTGTGTCTCT 517
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 565 TTTAGTCTGTGGTGGCTGGAGGCCACTCAAGNAGNAAACTGSCAAACTCTTAAAT 624
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 518 CTGAGTGCACAGAACTTGTGTAGATTCGCTCAACTGAAAAATATGGGAATAAAGCTGCAAT 577
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 625 CTGAGTCCCCAGAACCTAGTGGATTTCTGTCTCTCA-----GAATGATGCTGTGGA 675
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 GGGGTTTTCATGACAACTGCTTTCAGTATATTATTAACAAAGGCATTAATCAGAA 637
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 676 GGGGCTACATGACCAATGCTTCCATATATGTCAGAGAAGAACCGGGGTATTGACTCTGAA 735
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 GCTTCTATCCCTACAAAGCCATCAATGGAAAGTGCAGATATGACTCAAAAAACCGAGCT 697
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 GATGCTTACCCATATGTGGGACAGAGAGAGTGTATGTACAAACCCACAGGCAAGGCA 795
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 GCCACATGTTCAAAAGTATGTAAGTTCCTTTGGCAGTGAAGATGCTTAAAGAGCT 757
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 GCTAAATGCAGAGGTACAGAGATATCCCGAGGGGAATGAGAAGCCCTCAACAGGCA 855
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 GTGCCCAATAAAGGACCTGTGTGTGTGATAGATGAAATTAATTTCTTTCTGTG 817
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 GTGCCCGAGTGGGACCTGTCTGTGTGCCCATGTATGCAAGCTTGAACCTTCTTCTGAGTTT 915
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 818 TACAGAACTGTGTCTACTATGAACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 916 TACAGAAAGGTGTGTATTATGTAAGTGTGCAATAGCATAATCTGAACCATGGGTT 975
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 875 TTAGTGTGTGCTATGTTAAGCTTTAATGGGAAAGACTACTGGCTTGTGAAAAACAGCTGG 934
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 976 TTGCAGTGGATATGAAATCCAGAGGGAACAGCACTGGATTAATTAACACAGCTGG 1035
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 935 GGCCTCAACTTGTGTGACCAAGGATATATCGGATGCAAGCAACAGCTGGAATCACTGT 994
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1036 GGAGAAACTGGGAAACAAAGGATATATCTCATGGCTGGAATTAAGAAACAGCGCTGT 1095
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1096 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1129
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

83 CTGGACACCCACTGGGAGCTATGCAAGAACCCACAGAGCAATATAACAAGGTG 142
164 GAGGAAGTAGCAGCAGCTCTATCTCGGAGAAATACTCTAAAATTGTGATCTTCACAAT 223
143 GATGAATCTCTCGGCGTTTAAATTTGGGAAAAAACCTTGAAGTATATTTCCATCCATAAC 202
224 CTGGAACATCAATGGAATGATATCATATGATCTAGGCATGCAACCTCTGGGAGCATG 283
203 CTGAGGCTCTCTGCTGCTATCATATGATGATGATGATGATGATGATGATGATGATG 262
284 CTGAGGAGAGTATATCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 340
263 ACAGTGAAGAGGTGCTTCAGAGATGACTGCAAGTACCCCTGCTCATTTCCCGC 322
341 AGAATATGTCAC---TTATAGTCAAACTCTAATCAGAAATTCGCTGCTGTTGGGCT 397
323 AGTAATGACACCTTTATATCCAGAAATGGGAGGTAGAGCCCAAGACTCTGTCGACTAT 382
398 ACAGAGAGGGCTGTGCTTACTGATGAAATACCCAGGTTCTTTGGTCTGTTGGGCT 457
383 CGAAGAAAGGATATGTTACTCTCTCAAAATCAGGTCAGTGCTGTTGCTGTTGGGCT 442
458 TTATAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
443 TTTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
518 CTGAGTGCAGAACTTGGTAGATGCTCACTGAAATATGGGAAATGAGGCTGCAAT 577
503 CTGAGTGCAGAACTTGGTAGATGCTCACTGAAATATGGGAAATGAGGCTGCAAT 553
578 GCGGCTTTCATGACAACTGCTTCCAGTATATTTGATTAACAAACCGCATTTGATTCAGAA 637
554 GGGGCTCATGACCATGCTTCCATATGTCAGAGAACCGGGGTATGACTCTGAA 613
638 GCTTCTATCCCTACAAAGCCATGATGGAAGTGCAGATGACTCAAAAGAGCGCT 697
614 GATGCTTACCCATATGTTGGACAGAGAGAGTGTATGTATACAAACCCACAGCAAGGCA 673
698 GCCACATGTTCAAGTATCTGACTGAACTTCCCTTTGGCAGTCAACATCCCTTAAAGAGCT 757
674 GCTAAATGCAAGGGTACAGAGATCCCCCGAGGAATGAGAAACCCCTGCAAGCGCA 733
758 GTGCGGCTTACGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
734 GTGCGGCTTACGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
818 TACAGAGTGTGCTTACTTACTGAAACCATCTG---TACTCAGAAATGTGAATCATGGAGTA 874
794 TACAGCAAGGTGTCTATTATGATGAAGCTGCAATAGCGATATCTGAAACCATGCGGTT 853
875 TTAGTGTGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 934
854 TTGGCACTGGGATATGGAATCCAGAGGAAACAGCACTGGATTAATTAAGAAACAGCTGG 913
935 GCGCTCACTTTGCTGACCAAGGATATATTCGGATGCAAGAAACACTGGAATTAACATCTGT 994
914 GGAGAAACTGGGAAACAAAGGATATATCTCATGCTGCAATATAGAACAGCGCTGT 973
995 GGGATTTGCTTACTTCCCTCTTACCAGAAATCT 1028
974 GCGATTTGCAACCTGGCCAGCTTCCCAAGATGT 1007

```

RESULT 6

US-08-798-096-1
 ; Sequence 1, Application US/08798096
 ; Patent No. 6387682
 ; GENERAL INFORMATION:
 ; APPLICANT: Haslinds, et al.
 ; TITLE OF INVENTION: HUMAN OSTEOCALCIN-DERIVED CATHEPSIN
 ; FILE REFERENCE: PFI07D2
 ; CURRENT APPLICATION NUMBER: US/08/798,096
 ; CURRENT FILING DATE: 1997-02-12

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)..(1009)
; NAME/KEY: mat-peptide
; LOCATION: (365)..(1009)
; NAME/KEY: sig-peptide
; LOCATION: (21)..(1009)
US-08-798-096-1

```

Query Match 27.5%; Score 310; DB 4; Length 1619;
 Best Local Similarity 62.1%; Pred. No. 1.le-85;
 Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

```

QY 105 CTGGATCATCTACTGGAATCTCTGGAAGAAAACCTACAGCAACAAT-CAAGGAAGAGAAAT 163
DB 83 CTGGACACCCACTGGGAGCTATGGAAGAACCCACAGAGCAATATAACAAGAGTG 142
QY 164 GAGGAAGTAGCAGCAGCTCTATCTCGGAGAAATACTCTAAAATTGTGATCTTCACAAT 223
DB 143 GATGAATCTCTCGGCGTTTAAATTTGGGAAAAAACCTTGAAGTATATTTCCATCCATAAC 202
QY 224 CTGGAACATCAATGGAATGATATCATATGATCTAGGCATGCAACCTCTGGGAGCATG 283
DB 203 CTGAGGCTCTCTGCTGCTATCATATGATGATGATGATGATGATGATGATGATGATG 262
QY 284 ACTGTTGAAGAGTGTATCTTTGATGATGATGATGATGATGATGATGATGATGATGATG 340
DB 263 ACAGTGAAGAGGTGCTTCAGAGATGACTGCAAGTACCCCTGCTCATTTCCCGC 322
QY 341 AGAATATGTCAC---TTATAGTCAAACTCTAATCAGAAATTCGCTGCTGTTGGGCT 397
DB 323 AGTAATGACACCTTTATATCCAGAAATGGGAGGTAGAGCCCAAGACTCTGTCGACTAT 382
QY 398 ACAGAGAGGGCTGTGCTTACTGATGAAATACCCAGGTTCTTTGGTCTGTTGGGCT 457
DB 383 CGAAGAAAGGATATGTTACTCTCTCAAAATACAGGTCAGTGCTGTTGCTGTTGGGCT 442
QY 458 TTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
DB 443 TTTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
QY 518 CTGAGTGCAGAACTTGGTAGATGCTCACTGAAATATGGGAAATGAGGCTGCAAT 577
DB 503 CTGAGTGCAGAACTTGGTAGATGCTCACTGAAATATGGGAAATGAGGCTGCAAT 553
QY 578 GCGGCTTTCATGACAACTGCTTCCAGTATATTTGATTAACAAACCGCATTTGATTCAGAA 637
DB 554 GGGGCTCATGACCATGCTTCCATATGTCAGAGAACCGGGGTATGACTCTGAA 613
QY 638 GCTTCTATCCCTACAAAGCCATGATGGAAGTGCAGATGACTCAAAAGAGCGCT 697
DB 614 GATGCTTACCCATATGTTGGACAGAGAGTGTATGTATACAAACCCACAGCAAGGCA 673
QY 698 GCCACATGTTCAAGTATCTGACTGAACTTCCCTTTGGCAGTCAACATCCCTTAAAGAGCT 757
DB 674 GCTAAATGCAAGGGTACAGAGATCCCCCGAGGAATGAGAAACCCCTGCAAGCGCA 733
QY 758 GTGCGGCTTACGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
DB 734 GTGCGGCTTACGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
QY 818 TACAGAGTGTGCTTACTTACTGAAACCATCTG---TACTCAGAAATGTGAATCATGGAGTA 874
DB 794 TACAGCAAGGTGTCTATTATGATGAAGCTGCAATAGCGATATCTGAAACCATGCGGTT 853
QY 875 TTAGTGTGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 934

```

854	TTGGCATTGGGATATGTAATCTACGACGGGAAATACGCACTGGGATATTTAAACACAGCTGG	913
935	GGCTTCACCTTTGGTGACCAAGGATATATTTGGGATGGCAAGAAACAGTGGAAATCACTGT	994
914	GGAGAAAACCTGGGAAACCAAGGATATATCTCATGGCTGGAATTAAGAACCAACAGCCTGT	973
995	GGGATTGTATCTTATCTCTCTTTACCGAGAAATCT	1028
974	GGCATTTGCCAACCTGGGCAGCTTCCCAACATGT	1007

RESULT 7

US-08-798-095A-1

; Sequence 1, Application US/08798095A

; Patent No. 6423507

; GENERAL INFORMATION:

: APPLICANT: Hastings, et al.

7. AUTHOR: HUSKINS, ET AL.
: TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

FILE REFERENCE: PF107D3

US/08/798, 095A

; CURRENT FILING DATE: 1997-02-12

;	CURRENT FILING DATE:	199
;	NUMBER OF SEO ID NOS:	74

SOFTWARE: Patent In Ver. 2.1

: SEQ ID NO 1

: : SEQ ID NO: 1619
: : LENGTH: 1619

LENGTH: 1
TYPE: DNA

FILE: DNA
ORGANISM: Homo sapiens

: ORGANISM: HOMO SAPIENS
: FEATURE:

NAME/KFY - CDS

NAME/KEY: CDS
LOCATION: (21) (1003)LOCATION: (22) .. (100%)
NAME/KEY: mat peptide

```

; NAME/KEY: mac_peptide
; LOCATION: (365) (1009)

```

LOCATION: (365).. (1004)
NAME/KEY: sia pontido

```

; NAME./KEY: sig_peptide
; LOCATION: (21) (1009)

```

; LOCATION: (21
115-08-798-095A-1

Query Match

Best Local Similarity 62.18: Pred. No. 1.1e-85:

Best local similarity	92.1%	Freq. NO:	1.1e 05,
Matches	580;	Conservative	0; Mismatches
			335: Indels
			19: Gaps

[illegible]

PCT-US94-04781-1

Query Match 27.5% Score 310; DB 5; Length 1619;

Best Local Similarity 62.1% Pred. No. 1.1e-85;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGATCATCTGATCTCTGGAAGAAACCTCAGCAACAAT-CAAGGAAGAGAAAT 163
 DB 83 CTGACACCCACTGGGAGCTATCGAAGACACCCAGGAGCAATATAACAACAGGTG 142
 QY 164 GAGGAAGTACAGGGGCTCTCTGGAAGAAATCTAAATTTCTGATGCTTCAAT 223
 DB 143 GATCAATCTCTGGGCTTTAAATTTGGGAAAGAAACCTGAAGTATATTTCCATCCATAAC 202
 QY 224 CTGCAACACTCAATGGCAATCCATTCATATGATCTAGGCATGAACCATCTGGGAGACATG 283
 DB 203 CTGAGAGCTCTCTGTTGTTCCATACATATGAATGCTATGAACCACTGGGGAGACATG 262
 QY 284 ACTGGTGAAGAGTATATCTTTGATGGTTCCTGAGAGTTCC---AGCCAATGGCAG 340
 DB 263 ACCACTGAAGAGTGGCTTCAGAACTGACTGGACTCAAGTACCCCTGCTCATCTCCCGC 322
 QY 341 AGAATGTAC---TTATAGTCAAACTCTAATCAGAAATTTGCTGATCTGTGGACTGG 397
 DB 323 AGTAATGACACCTTTATATCCAGATGGGAAGGTAGAGCCCGAGACTCTGCGACTAT 382
 QY 398 AGAGAGAGGGCTGTGTTACTGAAATGAAATACCAGGGTCTTGTGTGCTTGTGGCT 457
 DB 383 CCAAGAAAGGATATCTTACTCTCTGCAAAATCAGGCTCAGTGTGTTCTGTGGCT 442
 QY 458 TTCAGGGTGTGAGGAGCTCTGGAAGCACTGAAGCTTAAACAGGAAAGCTGTGTCT 517
 DB 443 TTTAGCTGTGGGTGCTCTGGAGGCACTCAAGAGAACTGGAACAACTCTTAAAT 502
 QY 518 CTGAGTGCACAGAACTTGTGATGATGCTCAACTGAAATATGGAATAAGAGTGCAT 577
 DB 503 CTGAGTCCCAAGCTAGTGTGATGTTGTGTCTGA-----GAATGATGGGTGTGGA 553
 QY 578 GCGGTTCATCAGCACTCTCTTCCAGTATATTTATGATAACAGCGCATTTGATTCAGAA 637
 DB 554 GGGGCTACATGACCAATGCTTCCAAATATGTCAGAGAGAACCGGGGTATTGACTCTCAA 613
 QY 638 GTTCTCTATGCTTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAGAGCT 697
 DB 614 GATGCTTACCATATCTGGGACAGGAGAGTGTGATGATCAACCCCAACAGGCAAGGCA 673
 QY 698 GGCACATGTTCAAGTATCTGAATCTTCCCTTTGGCAGTGAAGATCCCTTAAAGAGCT 757
 DB 674 GCTAATGAGAGGGTACAGAGATCCCGGAGGGAATGACAAAGCCCTGAAGAGGCA 733
 QY 758 GTGGCCAAATAAGAGCTGTGTGCTGTGGCTATAGTGGCAGCCACTATTTCTTCTCTG 817
 DB 734 GTGGCCGAGTGGGAGCTGTCTGTGGCAATGATGCAAGGCTGACCTTCTTCCAGTTT 793
 QY 818 TACAGAGTGGTGTCTACTATGAACCATCTG---TACTCAGAAATGTGAATCATGGAGTA 874
 DB 794 TACAGAAAGGTGTGATATGATGAAGCTGCAATAGGCAATACTGAACCATGGGTT 853
 QY 875 TTAGTGTGTGGCTATGTAACCTTTAATGGGAAAGACTACTGGCTTGTGAAACAGCTGG 934
 DB 854 TTGGCAGTGGATATGAATCCAGAGGGAAGAAAGCACTGGGATAATTAAGAAACAGCTGG 913
 QY 935 GGCCTCAACTTTGTGACCAAGGATATTTCCGATGCAAGAAACAGTGGAAATCACTCT 994
 DB 914 GAGAGAAACTGGGAAAGAGGATATATCTCATGCTGCAATTAAGAACACAGCTCT 973
 QY 995 GAGATTGCTATTTATCTCTTTTACCCAGAAATCT 1028
 DB 974 GCAATTGCAACCTGGCAGCTTCCCAAGATGT 1007

RESULT 9

US-08-964-308-3

; Sequence 3, Application US/08964308

; Patent No. 6066715

; GENERAL INFORMATION:

; APPLICANT: DESMARAIS, SYLVIE

; APPLICANT: FRIESEN, RICHARD

; APPLICANT: ZAMBONI, ROBERT

; TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE

; TITLE OF INVENTION: BINDING ASSAY

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.

; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000

; CITY: RAHWAY

; STATE: NJ

; COUNTRY: USA

; ZIP: 07065

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Diskette

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/964,308

; FILING DATE: 04-NOV-1996

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: NORTH, ROBERT J

; REGISTRATION NUMBER: 27,366

; REFERENCE/DOCKET NUMBER: 19840 PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 732-594-7262

; TELEFAX: 732-594-4720

; TELEX:

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1669 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-964-308-3

Query Match 27.5% Score 310; DB 3; Length 1669;

Best Local Similarity 62.1% Pred. No. 1.1e-85;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGATCATCTGATCTCTGGAAGAAACCTCAGCAACAAT-CAAGGAAGAGAAAT 163
 DB 193 CTGACACCCACTGGGAGCTATGGAAGAAAGACCCAGGAGCAATATAACAACAGGTG 252
 QY 164 GAGGAAGTACGACGGCTCTCATCTGGGAAAGAAATCTAAATTTCTGATGCTTCAAT 223
 DB 253 GATCAATCTCTCGGCTTTAAATTTGGGAAAGAAACCTCAAGTATATTTCCATCCATAAC 312
 QY 224 CTGCAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGGAGACATG 283
 DB 313 CTGAGGCTCTCTTGTGTGCTCCATACATGAACTGGCTATGAACCACTGGGGACATG 372
 QY 284 ACTGCTCAAGAAAGTATATCTTTGATGGGTTCCTCAGAGTTGCC---AGCCAATGGCAG 340
 DB 373 ACCAGTGAAGAGTGGTTTCAGAAAGTGAAGTGAAGTCAAGTACCCCTGCTCATTTCCGCG 432
 QY 341 AGAATGTAC---TTATAGTCAAACTCTAATCAGAAATTTGCTGATCTGGGACTGG 397
 DB 433 AGTAATGACACCTTTTATATCCCAAGATGGGAAGGTAGACCCCACTCTGTCGACTAT 492
 QY 398 AGAGAGAGGGCTGTGTTACTGAAATAGCAAGGTTCTTTGTGTGTTGTTGTGCT 457
 DB 493 CGAAGAAAGGATATGTTACTCTCTCAAAATACAGGCTCAGTGTGTTGCTTGGGCT 552
 QY 458 TTCAGCGCTGTGGGGCCCTCGAAGACACACTGAAGCTAAACAGGAAAGCTGGTGTCT 517
 DB 553 TTAGCTCTGTGGTGGCTCTGAGAGAG---AA-TTAAAGAAAGTGAAGTGAAGT 612

```

QY 518 CTGAGTGCACAGAACTTGGTAGATTCTCAACTGAAATAATGGAATAAAGCGCTGCAAT 577
Db 613 CTGAGTGCACAGAACTGAGGATTGTGTGTGA-----GAATGATGGCTGTGGA 663
QY 578 GCGGGTTTCATGACAACTGCTTCCAGTATATATGATAACAAGCGGATTCATTCAGAA 637
Db 664 GGGGGCTACATGACCAATGCCCTCCAAATATGTCAGAGAACCGGGGTATTGACCTCGAA 723
QY 638 GTTCTCTATCCCTAANAAGCATGAATGGAAGTGCAGATATGATCAAAAAGGAGCT 597
Db 724 GATGCTTACCCATATGTGGCAGGAGAGAGATCTCTATGATAACCCCAAGGCAAGGCA 783
QY 698 GGCACATGTTCAAAGTATACTCAACTTCCCTTTGGCAGTGAAGATGCCCTTAAACAAGCT 757
Db 784 GCTAAATGACGAGGTACAGAGATGCCCGAGGGAATGAGAAGCCCTGAGAGGCA 843
QY 758 GTGGCCAAATAAGAGCTGTGTCTGTGGCTATAGATGGAGCCACTATCTTTCTTCCTG 817
Db 844 GTGGCCGAGTGGGAGCTGTGTCTGTGGCTATAGATGGAGCCACTATCTTTCTTCCTG 903
QY 818 TACAGAGTGGTCTCTACTATCAACCTCTG---TACTCAGAAATGTGAATCATGAGTA 874
Db 904 TACAGAAAGTGTGTATATGATAAGCTGCAATAGGATAATCTGAACCTCTCCAGTT 963
QY 875 TTAGTGGTGGTGTATGTAACCTTAATGGGAAGACTACTGCTTCTGAAAAACAGCTGG 934
Db 964 TTGGCAGTGGGATATGGAATCCAGAGGGAAGCAAGCACTGGATAATTAANAACAGCTGG 1023
QY 935 GGCCTCACTTTGGTGGCAAGGATATATTCGGATGGCAAGCAAGCTGGAATCAGTGT 994
Db 1024 GGAGAAAGCTGGGAAACAAAGGATATATCTTCATGCTCGAATTAAGAACCAAGCTGT 1083
QY 995 GGAATGCTAGTATATGCTCTTACCAGAAATCT 1028
Db 1084 GGCATGCCCACTGCCAGCTTCCCAAGATGT 1117

```

RESULT 10

```

US-08-964-313-3
: Sequence 3, Application US/08964313
: Patent No. 6114132
: GENERAL INFORMATION:
: APPLICANT: DESMARIS, SYLVIE
: APPLICANT: FRIESEN, RICHARD
: APPLICANT: GRESSER, MICHAEL
: APPLICANT: KENNEDY, BRIAN
: APPLICANT: NICHOLSON, DONALD
: APPLICANT: RAMACHANDRAN, CHIDAMBARAN
: APPLICANT: SKOREY, KATHRYN
: APPLICANT: FORD-HUTCHINSON, ANTHONY
: TITLE OF INVENTION: PHOSPHATASE BINDING ASSAY
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PHILIPPE L. DURETTE - MERCK & CO., INC.
: STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
: CITY: RAYWAY
: STATE: NJ
: COUNTRY: USA
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/964,313
: FILING DATE: 04-NOV-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/030,408
: FILING DATE: 04-NOV-1996
: APPLICATION NUMBER: PCT/CA97/00825

```

```

: FILING DATE: 03-NOV-1996
: AFTURNKEY/AGENT INFORMATION:
: NAME: DURETTE, PHILIPPE L.
: REGISTRATION NUMBER: 35,125
: REFERENCE/DOCKET NUMBER: 19824Y
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 732-594-4568
: TELEFAX: 732-594-4720
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1669 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-964-313-3

```

Query Match 27.5%; Score 310; DH 3; Length 1669;

Best Local Similarity 62.1%; Pred. No. 1,10-85;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

```

QY 105 CTGATATATCATCTGAAATCTTGGAGAAAACCTACAGCAAAATCAAGAGAGCAAT 163
Db 193 CTGCACACCCACTGGAGCTATGGAAGAACCCACAGAGCAATATAAACAACAGTG 252
QY 164 GAGGAAGTAGCAGCGCTCTCATCTGSAAMAAATATAAATTTGTGATCTTCATAT 223
Db 253 GATGAATCTCTGGCGTTTAATTTGGGAAAAAAGCTGAAGTAATTTCTATCCATAC 312
QY 224 CTGGAACACTCAATGGGAATGCAATTCATATCATCTAGGCATGAACATCTGGAGATG 283
Db 313 CTTCAGGCTTCTCTGGTGTATACATATGATGATGATGAACCTGAGAGATG 372
QY 284 ACTGCTGAAGAAAGTATATCTTTATGATGATGATGATGATGATGATGATGATG 340
Db 373 ACCAGTGAAGAGGTGGTTCAGAGATCATGGAGCTCAAGCTAGCCCTGCTCATTCGCC 412
QY 341 AGAAATGTCAC---TTATAGTCAAAATCTAAATTAATAATGATGATGATGATGATG 397
Db 433 AGTAATGAATAATTTATATATGATGATGATGATGATGATGATGATGATGATGATG 492
QY 398 AGACAGAGAGGCTGCTGTACTGAACTCAATATACCAGGCTTCTGCTGCTGCTGCTG 457
Db 493 GGAAGAGAAAGATATGTTATCTGTTAAATAGGCTAGTGTGATGATGATGATGATG 592
QY 458 TTAAAGCTGTGGGAGCTGGAAGCAAACTGAAGCTTAAATAAAGCAAGATGATGCT 517
Db 553 TTTAGCTCTGCTGGCTGCTGGAGGCTCACTCAAGAGAGAAAATGCAAACTCTTAAAT 612
QY 518 CTGATGATCAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 577
Db 613 CTGAGTGCACAGAACTGAGGATGATGATGATGATGATGATGATGATGATGATGATG 663
QY 578 GCGGTTTCATGACAACTGCTTCCAGTATATATGATAACAAGCGGATTCATTCAGAA 637
Db 664 GGGGGTACATGACCAATGCCCTCCAAATATGTCAGAGAACCGGGGTATTGACCTG 723
QY 638 GTTCTCTATCCCTAANAAGCATGAATGGAAGTGCAGATATGATCAAAAAGGAGCT 697
Db 724 GATGCTTACCCATATGTGGCAGGAGAGATCTCTATGATAACCCCAAGGCAAGGCA 783
QY 698 GGCACATGTTCAAAGTATACTCAACTTCCCTTTGGCAGTGAAGATGCCCTTAAACA 757
Db 784 GCTAAATGACGAGGTACAGAGATGCCCGAGGGAATGAGAAGCCCTGAGAGGCA 843
QY 758 GTGGCCAAATAAGAGCTGTGTCTGTGGCTATAGATGGAGCCACTATCTTTCTTC 817
Db 844 GTGGCTGAGATGGAAGCTGTGTCTGTGGCTATAGATGGAGCCACTATCTTTCTTC 903
QY 818 TACAGAGTGGTCTCTACTATCAACCTCTG---TACTCAGAAATGTGAATCATGAGTA 874
Db 904 TACAGAAAGTGTGTATATGATAAGCTGCAATAGGATAATCTGAACCTCTCCAGTT 963

```


398	ACACAGAAAGGCTGTGTACTTGAAGTGAATACACAGGGTCTCTGTGTGTGTCTTGGGCT	QY
400		
402		
404		
406		
408		
410		
412		
414		
416		
418		
420		
422		
424		
426		
428		
430		
432		
434		
436		
438		
440		
442		
444		
446		
448		
450		
452		
454		
456		
458	TTTACGGCTGTGGGGCCCTCGAAGCACAACTCAAGCTAAAAACAGGAAGCTGTGTGTCT	QY
460		
462		
464		
466		
468		
470		
472		
474		
476		
478		
480		
482		
484		
486		
488		
490		
492		
494		
496		
498		
500		
502		
504		
506		
508		
510		
512		
514		
516		
518	CTGAGTGCACAGAACTTGGTACATTGCTCAACTGAAAAATATCGGAAATAAGGCTCCAA	QY
520		
522		
524		
526		
528		
530		
532		
534	CTGAGTCCCCAGAACCTAGTGGATTGTGTGTGA-----GAATGATGGCTGTGGA	Db
536		
538		
540		
542		
544		
546		
548		
550		
552		
554		
556		
558		
560		
562		
564		
566		
568		
570	GGCGGTTTTCATGCAAAATGCTTTCAGTATATATTGATAACAAACGSCATTGATTCAGAA	QY
572		
574		
576		
578		
580		
582		
584		
586		
588		
590		
592		
594	GGGGCTACATGACCAATGCGCTTCCCAATATGTGCAGAGAACCGGGGTATTGACTCTGAA	Db
596		
598		
600		
602		
604		
606		
608		
610		
612		
614		
616		
618	GCTTCCTATCCCTACAAAGCCATCAATGGGAAGTGCAGATATGACATCAAAAAACCGAGCT	QY
620		
622		
624		
626		
628		
630		
632		
634		
636		
638		
640		
642		
644		
646		
648		
650		
652		
654	GATGCTACCCATATGTGGGACAGGAAGAGAGTTGTATGTACAAACCCACAGCGCAAGGCA	Db
656		
658		
660		
662		
664		
666		
668		
670		
672		
674		
676		
678		
680		
682		
684		
686		
688		
690		
692		
694		
696		
698	GCCACATGTTCAAAGTATACTGAACCTTCCCTTTGGCAGTGAAGATGCCTTAAAAAGAGCT	QY
700		
702		
704		
706		
708		
710		
712		
714		</

RESULT 15

```

RES-0011 13
US-08-964-308-13
Sequence 13, Application US/08964308
Patent No. 6066715
GENERAL INFORMATION:
APPLICANT: DESMARAIS, SYLVIE
APPLICANT: FRIMSEN, RICHARD
APPLICANT: ZAMHONI, ROBERT
TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
TITLE OF INVENTION: BINDING ASSAY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-BIOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,308
FILING DATE: 04-NOV-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

```

Db 835 TTGGCAGTGGGATATGGAATCCAGAGGGGAAACAGCACTGGGATAATTAAAAACACCTGG 894
QY 935 GGCTCAACTTTGGTGAACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCACTGT 994
Db 895 GGAGAAAACCTGGGGAAACAAAGGATATATCTTCATGGCTCGAAATAGAAACAACGCCCTGT 954
QY 995 GGGATTCCCTAGTTATCCCTCTTTACCCAGAAATCT 1028
Db 955 GGCATTGCTAACCTGGCTAGCTTCCCCAAGATCT 988

Search completed: July 1, 2003, 10:53:48
Job time : 81 secs

Db		241	AATGCATTATATGATCTAGCGATGAACCATCTGGGAGACAAGACCTGGTGAAGAAAGTCAT	300
Qy		301	ATCTTTGATGGGTCCCTGAGAGTTCCAGCCAAATGGCAGAGAAATGTCACTTATAGGTC	360
Db		301	ATCTTTGATGGGTCCCTGAGAGTTCCAGCCAAATGGCAGAGAAATGTCACTTATAGGTC	360
Qy		361	AAACTCTTAATCAGAAATTCGCTGATCTGTGTGGACTGGAGAGAGAAGCGCTGTCTACTGA	420
Db		361	AAACTCTAATCAGAAATTCGCTGATCTGTGTGGACTGGAGAGAGAAGCGCTGTCTACTGA	420
Qy		421	AGTGAANAATCCAGGGTCTTTGTGGTGCTTTGTGGGCTTTTCAGGCGTGTGGGGCCCTGGA	480
Db		421	AGTGAANAATCCAGGGTCTTTGTGGTGCTTTGTGGGCTTTTCAGGCGTGTGGGGCCCTGGA	480
Qy		481	AGCACAACTCAAGCTTAAACACAGAAAGCTGGTGTCTCTCAGTGCACAGAACTTGGTAGA	540
Db		481	AGCACAACTCAAGCTTAAACACAGAAAGCTGGTGTCTCTCAGTGCACAGAACTTGGTAGA	540
Qy		541	TTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTTCATGACAACTGCTTT	600
Db		541	TTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTTCATGACAACTGCTTT	600
Qy		601	CCAGTATATTATTGATAACACGGCATTTGATTCAGAAGCTTCCTATCCCTACAAAGCCAT	660
Db		601	CCAGTATATTATTGATAACACGGCATTTGATTCAGAAGCTTCCTATCCCTACAAAGCCAT	660
Qy		661	GAATGGAAAGTCGAGATATGACTCAAAAAAGCGAGTGCACATGTTTCAAGTATACTGA	720
Db		661	GAATGGAAAGTCGAGATATGACTCAAAAAAGCGAGTGCACATGTTTCAAGTATACTGA	720
Qy		721	ACTTCCCTTTGGCAGTGAAGATGCCTTTAAAGAAAGCTTGCCCAATAAAGAGCTGTGTC	780
Db		721	ACTTCCCTTTGGCAGTGAAGATGCCTTTAAAGAAAGCTTGCCCAATAAAGAGCTGTGTC	780
Qy		781	TGTGGCTATAGATGCGAGCCACATATCTTTCTTCCTGTACAGAAGTGGTGTCTACTATGA	840
Db		781	TGTGGCTATAGATGCGAGCCACATATCTTTCTTCCTGTACAGAAGTGGTGTCTACTATGA	840
Qy		841	ACCATCCTGTACTCAGAAATGGAATCATGGAGTATTAGTGTTGGCTATGGTAAACCTTAA	900
Db		841	ACCATCCTGTACTCAGAAATGGAATCATGGAGTATTAGTGTTGGCTATGGTAAACCTTAA	900
Qy		901	TGGGAAGACTACTGGCTTGTGAAAACAGCTGGGGCCCTCACTTTGGTGACCAAGGATA	960
Db		901	TGGGAAGACTACTGGCTTGTGAAAACAGCTGGGGCCCTCACTTTGGTGACCAAGGATA	960
Qy		961	TATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATTCGTAGTTATCCCTCTTACCC	1020
Db		961	TATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATTCGTAGTTATCCCTCTTACCC	1020
Qy		1021	AGAAATCTAGACCTCTTCATTTTATACAAAGTCCAAAAAATTCAAACACTTTCTCTTAAT	1080
Db		1021	AGAAATCTAGACCTCTTCATTTTATACAAAGTCCAAAAAATTCAAACACTTTCTCTTAAT	1080
Qy		1081	TTAATTTTACCTGCTATAATAGTAAAAAATAAATGTGTCAATGACCAT	1126
Db		1081	TTAATTTTACCTGCTATAATAGTAAAAAATAAATGTGTCAATGACCAT	1126

RESULT 2

```

RESOLUTION 2
US-10-099-275-1
: Sequence 1, Application US/10099275
: Publication No. US20020187499A1
: GENERAL INFORMATION:
: APPLICANT: Schneider, Patrick
: APPLICANT: Yamamoto, Karen K.
: APPLICANT: French, Cynthia K.
: APPLICANT: Reprogen, Inc.
: TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
: TITLE OF INVENTION: Endometriosis
: FILE REFERENCE: 018002-00131005
: CURRENT APPLICATION NUMBER: US/10/099,275
: CURRENT FILING DATE: 2002-04-13

```



```

      812 TGTCTGTTGGTGTACATGCGGCTCACTCTTCTTCTCTACAGAAAGTGGTGTCTACT 871
      QY
      837 ATGAACCATCTCTACTCAGAAATGTGAATCATGGATGATTAATAGTGGTATGGTAAACC 896
      Db
      872 ATGAACCATCTCTACTCAGAAATGTGAATCATGGTGTACTGTTGGTGTGGTATGGTGTATC 931
      QY
      897 TTAATGGGAAAGAACTACTGGCTTCTGAAAGAACAGCTGGGCGCTCAACTTTTGGTGTGACCAAG 956
      Db
      932 TTAATGGGAAAGAACTACTGGCTTCTGAAAGAACAGCTGGGCGCTCAACTTTTGGTGTGAAAG 991
      QY
      957 CATATATTCGGATCGCAAGAACAGTGGAAATCACTGTGGGATTTAGTATTCCTCTTT 1016
      Db
      992 CATATATTCGGATCGCAAGAAATTAAGGAAATCACTGTGGGATTTAGTATTCCTCTTT 1051
      QY
      1017 ACCGAAATCTA---GACCTCTTCACTTTTATAAAGTCCCAAAATATGAAGCACTTT 1072
      Db
      1052 ACCGAAATCTAAGGATCTCTCTTTTATAAATCAATGAATATGAAGCACTTT 1111
      QY
      1073 CTCTTAATTTAATTTTACTGCTATA-ATAGTAAATTAATTTGTCATGA 1122
      Db
      1112 CTCTTAATTTAATTTTCTGCTGTATCCAGAAAGAAATTAATTTGTCATGA 1162

RESULT 3
US-09-990-064-1
; Sequence 1, Application US/09990064
; Patent No. US20020164765A1
; GENERAL INFORMATION:
; APPLICANT: Parvinder K. Punia
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30242
; CURRENT APPLICATION NUMBER: US/09/990, 064
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: GB028462.0
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Mammalia
US-09-990-064-1

```

```

Query Match      69.3%; Score 784; DB 9; Length 996;
Best Local Similarity 87.3%; Pred. No. 1.4e-222;
Matches 870; Conservative 0; Mismatches 125; Indels 1; Gaps 1;

QY 36 ATGAATGGCTAGTTGGGCTGCTCCGTTGCTCTATGCACGACCACTACATAAA 95
Db 1 ATGAATGGCTAGTTGGGCTGCTCCGTTGCTCTATGCACGACCACTACATAAA 60
QY 96 GATCCCACTCTGGATCATCACTGGAATCTCTGGAAGAAACCTACAGCAAAACAT-CAAG 154
Db 61 GACCTACCTGGATCACTGACCTGCTCTCTGGAAGAAACCTACAGCAAAACATCAAG 120
QY 155 GAAGAGAAATGAGAAATGAGCAAGGCTCTCATCTGGGAAAGAAATCTAAATTTGTGATC 214
Db 121 GAAAGAAATGAGAAAGAGCAAGGCTCTCATCTGGGAAAGAAATCTGAGTTGTGAGC 180
QY 215 CTTCACATCTGGAACACTCAATGGGAATCATTCATATGATAGGCAATGAACCACTG 274
Db 181 CTTCACAACTGGGAGCACTCTCATGGGAATGACATCTGATGACGTGGCATGAACCACTG 240
QY 275 GCAACATGATGATGAGAAAGTATCTTTGATGGGTTCCCTGAGAGTTCCAGGCAA 334
Db 241 GCAGACATGACCAAGCAAGAAAGTGGTGTCTTTGATGATGATGATGATGATGATGATG 300
QY 335 TGGGAGAAATGTCATTTATAGTCAAACTCTAATCAGAAATTTGCTGATTTCTGTGGAC 394
Db 301 TGGCGGAGAAAGCTCACTTACAAAGTTAAACCCCAATCAGAAATTTGCTGATTTCTGTGGAC 360
QY 395 TCGAGAGAGAGGCGCTGTGTTACTGAAGTGAATACCAAGGCTTCTGTGCTGTGTTGG 454

```

```

      361 TCGAGAGAGAGGGGTGTGTCACATGAAGTCAAAATATACGGTCTTCTGTGCTGCTGG 420
      QY
      455 GCTTTTACGCGCTGTGGGGCCCTGGAAAGCAAACTGAAGTCAAAACAGCAAAAGCTGGTG 514
      Db
      421 GCTTTCAGTGGCGTGGGGCTCTGGAAGCACAGCTGAAGCTGAAGAGCAAAATCTGGTG 480
      QY
      515 TCTCTGAGTGCACAGAACTTGGTATGATTCACCTCAAAATATATGGGAATAAAGCTGC 574
      Db
      481 TCTCTGAGTGCACAGAACTTGGTATGATTCACCTCAAAATATATGGGAATAAAGCTGC 540
      QY
      575 AATGCGGTTTTCATCACAACCTGCTTCCAGTATATTATTCATAAACAAGGCTATGATTC 634
      Db
      541 AATGCGGTTTTCATCACAAGAGGCTTCCAAATACATCATGATCAACAACGCGCATCGATTC 600
      QY
      635 GAAGCTTCTATCCCTACAAAGGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 694
      Db
      601 GAAGCTTCTATCCCTACAAAGGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 660
      QY
      695 GCTGCGCATGTTTCAAGTATATCACTTCCCTTTTCCAGTGAAGTGAAGTGAAGTGAAG 754
      Db
      661 GCTGCGCATGTTTCAAGTATATCACTTCCCTTTTCCAGTGAAGTGAAGTGAAGTGAAG 720
      QY
      755 GCTGCGCATGTTTCAAGTATATCACTTCCCTTTTCCAGTGAAGTGAAGTGAAGTGAAG 814
      Db
      721 GCTGCGCATGTTTCAAGTATATCACTTCCCTTTTCCAGTGAAGTGAAGTGAAGTGAAG 780
      QY
      815 CTGTACAGAAAGTGGTGTCTACTATGAAGCACTCTCTACTCAGAAATGTAATCATGGAGTA 874
      Db
      781 CTGTACAGAAAGTGGTGTCTACTATGAAGCACTCTCTACTCAGAAATGTAATCATGGAGTA 840
      QY
      875 TTAGTGGTGGCTATGGTAACTTAAATGGGAAGCACTACTGCTTGTCAAAACAGCTGG 934
      Db
      841 CTAGCAGTGGCTATGGTAACTTAAATGGGAAGCACTACTGCTTGTCAAAACAGCTGG 900
      QY
      935 GGCTTCAACTTTGGTGACCAAGATATATTCGGATGCGCAAGAAACAGTGGAAATCATG 994
      Db
      901 GGCATCACTTGGTGAACAGATATATTCGGATGCGCAAGAAACAGTGGAAATCATG 960
      QY
      995 GGGATGCTAGTTATCCCTTCTACCCAGAAATCTAG 1030
      Db
      961 GGGATGCTAGTTATCCCTTCTACCCAGAAATCTAG 996

```

```

RESULT 4
US-10-010-577-1
; Sequence 1, Application US/10010577
; Publication No. US20030104971A1
; GENERAL INFORMATION:
; APPLICANT: thurmond, robin l
; APPLICANT: baker, sherry
; APPLICANT: karlsson, lars
; TITLE OF INVENTION: polynucleotide and polypeptide sequences of monkey cathepsin s
; FILE REFERENCE: ORT1457
; CURRENT APPLICATION NUMBER: US/10/010,577
; CURRENT FILING DATE: 2001-11-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 1
; LENGTH: 993
; TYPE: DNA
; ORGANISM: monkey
US-10-010-577-1

```

```

Query Match      68.2%; Score 768; DB 9; Length 993;
Best Local Similarity 87.0%; Pred. No. 7.9e-218;
Matches 867; Conservative 0; Mismatches 125; Indels 4; Gaps 2;

QY 36 ATGAATGGCTAGTTGGGCTGCTCCGTTGCTCTATGATGATGCGACCAAGTACATAAA 95
Db 1 ATGAAGCAGCTGGTTGTGTGCTGTTTGTGCTCTCTGCGGTGACACAGTTGCATATA 60
QY 96 GATCCCACTCTCGATCATCACTGGAATCTCTGGAAGAAACCTACAGCAAAACAT-CAAG 154

```

```
Db 61 GATCCACCCCTGGATCACCACCTGGAATCTCGAAGAAAAACCTACGGCAACAATATAA 120
QY 155 GAAGACAAATGAGGAAGTACGACGGGTCTCATCTCGGAAAAAATCTAAATTTGTGATG 214
Db 121 GAAAGAAATGAAGACGATAGACGGTCTCATCTCGGAGAGAAATCTAAAGTTTGTGATG 180
QY 215 CTTCAACAATCTGGAACACTCAATGGATGCAATTCATATGATCTAGGCATGAACCATCTG 274
Db 181 CTTCAACAATCTGGAACACTCAATGGATGCAATTCATATGATCTAGGCATGAACCATCTG 240
QY 275 GGAGACATGACTGGGTGAGGAAGTATATCTTTGAGGTTCCCTCGAGAGTTCCTCCAGCCAA 334
Db 241 GGAGACATGACTGGGTGAGGAAGTATATCTTTGAGGTTCCCTCGAGAGTTCCTCCAGCCAA 300
QY 335 TGGCAGAGAAATGTCACCTTATAGGTCACAACTCTAATCAAGATTCCTGATCTGTGCAC 394
Db 301 TGGCAGAGAAATGTCACCTTATAGGTCACAACTCTAATCAAGATTCCTGATCTGTGCAC 360
QY 395 TGGAGAGAGAGGGGTGTGTACTGAAAGTGAATACCAAGGTTCTTTGGTGGTCTTTGG 454
Db 361 TGGAGAGAGAGGGGTGTGTACTGAAAGTGAATACCAAGGTTCTTTGGTGGTCTTTGG 420
QY 455 GCTTTCAGGCTGTGGGGCCCTGGAAGCACAACCTGAAGCTGAAGCTGAAGCTGAAGCTGA 514
Db 421 GCTTTCAGGCTGTGGGGCCCTGGAAGCACAACCTGAAGCTGAAGCTGAAGCTGAAGCTGA 480
QY 515 TCTCTCAGTGGCAGACAACTTGGTACATTCCTCAACTGAAATATGGAATTAAGGCTGC 574
Db 481 TCTCTCAGTGGCAGACAACTTGGTACATTCCTCAACTGAAATATGGAATTAAGGCTGC 537
QY 575 AATGGGGTTTATATCAAAATGCTTTTCCAGTATATTTGATACCAAGGCTGAATGCA 634
Db 538 AATGGGGTTTATATCAAAATGCTTTTCCAGTATATTTGATACCAAGGCTGAATGCA 597
QY 635 GAAGCTTCTATCCCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAAGCA 694
Db 598 GAAGCTTCTATCCCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAAGCA 657
QY 695 GCTGCCACATGTTCAAAAGTATACCTGCTTCTGAGTGAAGTGCCTTAAAGAA 754
Db 658 GCTGCCACATGTTCAAAAGTATACCTGCTTCTGAGTGAAGTGCCTTAAAGAA 717
QY 755 GCTGTGGCAATTAAGGACCTGTGTGTGGCTATAGATGCGAGCCACTATCTTCTC 814
Db 718 GCTGTGGCAATTAAGGACCTGTGTGTGGCTATAGATGCGAGCCACTATCTTCTC 777
QY 815 CTGTACAGAGTGGTGTCTACTATGAACCATCTCTACTCAGAATGTGAATCATGGAGTA 874
Db 778 CTGTACAGAGTGGTGTCTACTATGAACCATCTCTACTCAGAATGTGAATCATGGAGTA 837
QY 875 TTAGTGGTTGGCTATGGTTAACTTAAATGGGAAGCACTACTGGCTTGTGAAAAACAGCTGG 934
Db 838 CTGTGATGGCTATGGTTAACTTAAATGGGAAGCACTACTGGCTTGTGAAAAACAGCTGG 897
QY 935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGAATCACTGT 994
Db 898 GGCACCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAATTAAGGAAACCACTGT 957
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCTAG 1030
Db 958 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCTAG 993
```

RESULT 5

```
US-09-918-995-36579
; Sequence 36579, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36579
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36579
```

```
Query Match 29.6%; Score 331.4; DB 9; Length 417;
Best Local Similarity 87.5%; Pred. No. 3.5e-88;
Matches 365; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
```

```
QY 393 ACTGGAGAGACAAGGGTGTGTACTGAAGTGAATACCAAGGTTCTTTGGTGGTCTTTGGT 452
Db 1 ACTGGAGAGACAAGGGTGTGTACTGAAGTGAATATCAAGTCTTTTGGTGGTCTTTGGT 60
QY 453 GGGCTTTTCAGCGCTGTGGGGCCCTGCAAGCACAACCTGAAGCTAAAAACAGCAAGCTGG 512
Db 61 GGGCTTTTCAGTGTGTGGGGCCCTGCAAGCACAACCTGAAGCTAAAAACAGCAAGCTGG 120
QY 513 TGTCTCTGAGTGTCAAGAACTTGTGATGATTTCTTAATGAAAAATATGGAATATAAGCT 572
Db 121 TGTCTCTCAGTGTCCCAACCACTGCTGATTTGCTCAACTGAAAAATATGGAATATAAG 180
QY 573 GCAATGGGGTTTCATCACAACCTGCTTCCAGTATATTAATGATAACAGGCTATTCAT 642
Db 181 GCAATGGTGTCTTCAATGCAAGGCTTTCCAGTATATTAATGATAACAGGCTATTCAT 240
QY 633 CAGAAGCTTCTATCCCTACAAAGCCATGAATGCAAGTGCAGATATGATTTAAAAAGC 692
Db 241 CAGAGCTTCTATCCCTACAAAGCCATGCATCAGAAATGCAATATGACTCAAAATATC 300
QY 693 GAGTGGCCACATGTTCAAGTATATCTGAACTTCTGCTTTGGCACTGGAAGTGGCTTAAAG 752
Db 301 GTGCTGCCACATGTTCAAGTATATCTGAACTTCTGCTTTGGCACTGGAAGTGGCTTAAAG 360
QY 753 AAGCTGTGGCAATTAAGGACCTGTGTGTGGCTATAGATGCGAGCCACTATCTTCT 804
Db 361 AAGCTGTGTGTATTAAGGCTGTGTGTGGCTATAGATGCGAGCCACTATCTTCT 417
```

RESULT 6

```
US-09-918-995-37273
; Sequence 37273, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37273
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-37273
```

```
Query Match 29.4%; Score 331.4; DB 9; Length 418;
Best Local Similarity 87.7%; Pred. No. 3.5e-88;
Matches 362; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 520 GAGTGGACAGAACTTGTGATGATTTGCTCAACTGAAAAATATGGAATATAAGGCTGAATGC 574
Db 6 GCGTGGCAGAACTTGTGATGATTTGCTCAACTGAAAAATATGGAATATAAGGCTGAATGC 65
```


US-10-198-846-10186

Query Match 27.5%; Score 310; DB 9; Length 2017;
Best Local Similarity 62.1%; Pred. No. 2.2e-81;

QY	105	CTGGATCATCACTCGAATCTCTCGAAGAAAACCTACAGCAACAAT-CAAGGAAGACAAT	163
Db			
Db	192	CTGGACACCCACTCGGAGCTATGGAAGAAGAGCCACAGAGCAATATAACCAACAGGTG	251
QY	164	GAGGAAGTAGCACGGCTCTCATCTCGGAAAAAAATCTAAAAATTTGTGATCTTCACAAT	223
Db			
Db	252	GATGAATCTCTCGGCGTTAATTTGGGAAAAAAACCTGAAGTATATTTCCATCCATAAC	311
QY	224	CTGGAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGGAGACATG	283
Db			
Db	312	CTTGAGGCTTCTCTTGCTGCCATACATATGAACCTGGCTATGAACCACTCTGGGGACATG	371
QY	284	ACTGCTGAAGAAGTGATATCTTTTGATGGTTCCTTCAGAGTTCCG- --AGCCAATGGCAG	340
Db			
Db	372	ACCAGTGNAAAGTGTGTTTCAGAAGATCAATTAAGTAACTTCTGCTCTATTCCCGC	431
QY	341	AGAAATGTCAG- --TTATAGGTCAAACTCTAATCAGAAATTCGCTGATTCTGTGAGCTGG	397
Db			
Db	432	ACTAATGACACCCCTTTATATCCAGAAATGGGAAGGTAGAGCCCACTCTGTCCGACTAT	491
QY	398	AGAGAGAAGGCTGTGTTACTGAAGTGAATACCAAGGGTCTTGTGGTGTCTTGTGGGCT	457
Db			
Db	492	CGAAGAAGAAGATATGTTACTCTCTGCAAAAATCAGGTCAGTGTGGTCTCTGTGGGCT	551
QY	458	TTACGCCCTGTGGGGGCCCTGGGAAGCACAACTCAAGCTAAAAACAGGAAAGCTGGTCTCT	517
Db			
Db	552	TTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAGAAAACTGCGAAACTCTTAAAT	611
QY	518	CTGAGTCGACAGAACTTGTGTAGATTCCTCAACTGAAAAATATGGGAATTAAGAGCTGCAAT	577
Db			
Db	612	CTGAGTCCCAGAACCCTAGTGGATGTGTGTCTGA- -----GAATGATGGGTGTGGA	662
QY	578	GGGGTTTCATGCAACTGTCTTCCAGTATATTTGATTAACAACTGCAACGCAATGATTCAGAA	637
Db			
Db	663	GGGGGCTACATGACCAATGCTTCCAAATATGTCCACAAGACCGGGGTATTTGACTCTGAA	722
QY	638	GCTTCCCTATCCCTACAAGGCCATGAATGGAAGCTGCAGATATGACTCAAAAACCGGAGCT	697
Db			
Db	723	GATGCTTACCTCTATATGTGGAAGAAATATGATAT- --AAACCAAAGCAAGGCA	782
QY	698	GCCACATGTTCAAGTATACTGAACCTTCCCTTTGGCAGTGAAGATGCCCTTAAAAAGAGCT	757
Db			
Db	783	GCTAAATGCCAGGGTACAGAGATCCCCGAGGGGAATGAGAAAGCCCTCAACAGGGCA	842
QY	758	GTGGGCAATAAAGAAATGT	817
Db			
Db	843	GTGGCCCGAGTGGGAGCTGTCTGTGGCACTTGATGCAAGCCCTGACCTCTCTCCAGTTT	902
QY	818	TACAGAAGTGTGTCTACTATGAACCATCTCG- --TACTCAGAATGTCAATCATGGAGTA	874
Db			
Db	903	TACAGCAAGGTGTGTTATGATGAAGCTCCATAGCGATATCTGACCATTCGCGTT	962
QY	875	TTAGTGTGTGCTATGTTAACCCTTAATGGGAAGACTACTGGCTTGTGAAAAACAGCTGG	934
Db			
Db	963	TTGGCACTGGGATATGGAATCCAGAAGGGAAAAACAGCACTGGATTAATTAATAACAGCTGG	1022
QY	935	GGCTCTCAACTTGTGTACCAAGCATATATTCGGATGGGCAAGAAACAGTGGGAATCACTGT	994
Db			
Db	1023	CGACAAAACCTGGGAAACAAGGATATATCTTCATGCGCTCGCAATTAAGAACACAGCCCTGT	1082
QY	995	GGGATTCTAGTTATCCCTCTTACCCAGAAATCT	1028
Db			
Db	1083	GGCAATGCCAAGCTTGGCGAGTCTTCCGCAAGATGT	1116

RESULT 12

```

US-09-960-352-13717
; Sequence 13717, Application US/09960452
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Hyatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10294)C
; CURRENT APPLICATION NUMBER: US/09/960,452
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13717
; LENGTH: 329
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 59-HOVMS1-011-Q1-E1-G12
; US-09-960-352-13717

```

Query Match	24.1%	Score	271.4;	DB	10;	Length	329;
Best Local Similarity	89.1%;	Pred No.	2.1e-70;				
Matches	293;	Conservative	0;	Mismatches	36;	Indels	0;
Gaps	0;						
<hr/>							
QY	665	GCAAAGTCCAGATATGACTCAAAAAGGGACCTGCCACATGTTCAAGGTATATCTGAATT	724				
Db	1	GGAAAGTGCATATGATGATATAAANAATCGATTCGCAAGTGTTCAAGGTATATCTGAATT	60				
QY	725	CCCTTTGGCAGTGAAGATGCCCTTAAAAAGAAGCTTGCCCAAIAAAGCACCTCTGTTCTTG	784				
Db	61	CCCTTCGGCGGAGAGAGGCCCTTAAAGAGAGCCGTCGCCCAATAAGAGCACCTGTTCTGTT	120				
QY	785	GGTATAGATGGAGTGCACCTATTCTTTTCTCTATACAGAGTGGTGTCTACTATGAACCA	844				
Db	121	GGTATAGAGCGGACGCCATTCTTCTTCTCTCTACAAAACGGTCTCTACTATCAACCCC	180				
QY	845	TCCGTACTACAATGTGAATCATGGAATTAAGTGTGGCTATGCTAACCTTAAATGGG	904				
Db	181	TCTCTAATACAAATGTAAACATGTATATCTGTGGTGGTATGATGATCTGAACGGG	240				
QY	905	AAAGACTACTGGCTGTGTGAAAAACAGTGGGGGCTCAAGCTTGGTCACCAACATATATT	964				
Db	241	AAGGACTACTGGCTGTGTGAAAAACAGCTGGGGGCTGCACCTTGGTCACCAACATATATA	300				
QY	965	CUGATGGCAAGAAACASTGSAATCACTG	993				
Db	301	CGCATGGCAAGAAATAGTGGAAATCACTG	329				

```

RESULT 13
US-09-960-352-12526
; Sequence 12526, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12526
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB34-084-Q1-E1-F2
US-09-960-352-12526
Query Match 22.3% Score 250.8; DB 10; Length 369;

```


Query Match	21.2%	Score 239.2	DB 9	length 1342
Best Local Similarity	57.3%	Pred. No. 2e-60		
Matches 539	Conservative 0	Mismatches 378	Indels 23	Gaps 5
Qy	96	GATCCCCACTCTGGATCATCTAC	TGCAATCTCTGGAGAAACCTACAGCAACAATCAAGG	155
Db	123	GACCAAAATTGGATACAAACTG	TACCACTGGAAAGGCACACACAG--AAGATTATATG	180
Qy	156	AAGAGATAGGAAATAGTACAGG	GGTCTCATCTGGGAAAAAATCTAAAAATTTGTGATGC	215
Db	181	CGCGGAATGAAGAGATGGAGG	AGACAGTGTGGAAAGAAATATCAAAATGATTCGAAC	240
Qy	216	TTCACAATCTGGAAACATCAAT	GGAAATGCATATCATATGATCTAGGCATGAACCATCTGG	275
Db	241	TGCNAATGGGAAATACAGCC	AGGGAACATGCTTCACAAATGGCCATGAATGCTTTTG	300
Qy	276	GAGACATGACTGGTGAAGAA	GTATATCTTTGATGGGTTCCTTGAGAGTTCCCAAGCCAAAT	335
Db	301	GTGACATGACCAATGAAGAA	ATTCAGGCAGATGATGGCTTGCTTCGAAACACAGAAATTC	360
Qy	336	GGCAGAGAAATGTCATATAG	GTCAAACTCTATACAGAAATTCGCTGATCTGTGCACT	395
Db	361	GGAAGGGAAGTGTCCGTG	ACCTCTGT--TTCCTTGAATCTCCCAATCTGTGGATT	417
Qy	396	GGAGAGAGAAGGCTGTGTT	ACTGAAGTGAATACCAAGGGTCTCTGTGGTCTTTGTTGG	455
Db	418	GGAGAAAGAAGGCTAGCTG	CAGCCAGTGAAGAAATCAGAAACAGTGTGTTCTTGTGGG	477
Qy	456	CTTTACGCGTGTGGGGCCCT	GGACGACCAACTGAAGCTAAAAACAGGAAGCTGCTGT	515
Db	478	CTTTAGTGGACTGTGTCCT	TGAAGCAGATGTTCCGGAACCTTGGGAACCTGTCT	537
Qy	516	CTCTGAGTCACAGAACTT	GGTAGTTGCTCAACTGAAAAATATGGAAATAAAGGCTGCA	575
Db	538	CACTCGCGAGCAAACTCT	GGTGCATGTTTCGGGTCTCTCAA--GGCAATCAGGGCTGCA	594
Qy	576	ATGCGCTTTTCATCACAACT	GCTTTCCAGTATATATTGATTAACAACGGGATTCATTTCAG	635
Db	595	ATGGTGCCTTCATGGCT	TAGGCGCTTCAGTATGTCAAGGAGACGGAGGCTGGACTCTG	654
Qy	636	AAGTTCTCTATCCCTACA	AAAGCCATGAATGGAAAGTCAGATATGACTCAAAAAGCGAG	695
Db	655	AGAAATCCTATCCATATG	TAGCAGTGGATGAAATCTGTAAGTACAGACCTCAGAAATCTG	714
Qy	696	CTGTCACATCTTCAAACT	ATATCTCAACTTCCTTTTGGCAGTGAAGATGCTTAAACAAG	755
Db	715	TTGCTAATGACATGGCT	TCACAGTGTCCGACCTGGAAAGGACAGGCGCTGATGAAG	774
Qy	756	CTGTGGCCAAATAAGAC	ACCTGTGTCTGGGTATAGATGGAGGCCATATCTTTTCTTCC	815
Db	775	CAGTCGCAACTGTGGG	CCCATCTCCGTTGTCTATGGATGCAAGGCCATCTCGCTTCCAGT	834
Qy	816	TGTACACAAGTGTGTCT	ACTATCAACCA--TCCTGTACTCAAAATGTCAATCATCGAG	872
Db	835	TCATACAAATCAGGCA	TTTATTTTCAACCACACTGCACACAGCAAAACCTGGATCATGGT	894
Qy	873	TATTAGTGGTTGGCTAT	GTG-----TAACCTTAATGGGAAAGACTACTGSGCTTG	920
Db	895	TTCTGTGTGTGGCT	ACGGCTTTCAAGGAGCAAAATTCGAATAACAGCAAGTATTGGCTCG	954
Qy	921	TGAAAAACACTGGGGCC	TCAACTTTTGGTCACCAAGGATATATTCCGATGGCAGAAACA	980
Db	955	TCAAAACACAGTGGG	TCAGATAGGCTATGTAAATAATAGCCAAAGACA	1014
Qy	981	GTGGAATCACTGTGG	GAATGCTATACCTTCTTACCC	1020
Db	1015	AGAACCAACCTGTG	CAATGCCACAGCAGCCAGCTACCC	1054

Search completed: July 1, 2003, 12:03:59
Job time : 232 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:22:27 ; Search time 2955 Seconds
(without alignments)
9580.522 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1126
Sequence: 1 gcaattatgtatgagagca.....aataatgtgtcatgacct 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Pending_Patents_NA_Main..*

```

1: /cgn2_6/ptodata/1/pna/US06_FCTUS_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq.*

```

```

44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq.*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1126	100.0	1126	38	US-10-010-580-1
2	856.4	76.1	2205	23	US-09-606-680-4112
3	856.4	76.1	2205	25	US-09-652-123-9177
4	856.4	76.1	2205	25	US-09-652-126-9929
5	856.4	76.1	2205	28	US-09-710-281-4558
6	856.4	76.1	2205	28	US-09-716-972-4347
7	856.4	76.1	2205	28	US-09-716-972-4347
8	856.4	76.1	2205	29	US-09-721-589-5798
9	856.4	76.1	2205	29	US-09-726-172-2249
10	856.4	76.1	2205	29	US-09-726-788-6295
11	856.4	76.1	2205	29	US-09-726-802-2191
12	856.4	76.1	2205	29	US-09-726-804-1915
13	856.4	76.1	2205	29	US-09-726-811-5142
14	856.4	76.0	2205	30	US-09-760-475-1251
15	856.4	76.0	2205	30	US-09-760-475-1251
16	854.8	75.9	1837	17	US-09-359-922-8469
17	854.8	75.9	1837	17	US-09-359-922-8469
18	854.8	75.9	1837	34	US-09-919-002-8469
19	853.2	75.8	1848	76	US-60-324-185-4019
20	847.2	75.2	2216	65	US-60-214-360-3680
21	845.4	75.1	1643	39	US-10-090-275-1

Sequence 1, Appli


```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4312
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2205)
; OTHER INFORMATION: n = A,T,C or G
; US-09-606-680-4312

```

```

Query Match          76.1%; Score 856.4; DB 23; Length 2205;
Best local Similarity 87.3%; Pred. No. 1.8e-226;
Matches 986; Conservative 0; Mismatches 136; Indels 8; Gaps 4;

```

QY	1	GCATATTAGTAGTGGAGCACCTCCAGGTTCTTTTC---	ATGAAATGCTAGTTGGGCTGC	57
DB	222	GGACHTCTTCTCTGGGAGCAACTCGTGTCTATCA	CAATGAAAGGCCTGGTTGTGTGC	281
QY	58	TTCCCTTTGCTCTCTATGCACTGGCACAAGTACATA	AAAGATCCCACTCTCGATCATCACT	117
DB	282	TCTTGGTGTGCTCTCTCGACTGGCAGTTGCAATA	AAAGATCCTTCCCTGGATCACCACCT	341
QY	118	GCAATCTCTGGAAGAAAACCTACAGCAACAAT--	CAAGGAAGAGAAATGAGGAAGTAGCAC	176
DB	342	GGCATCTCTGGAAGAAAACCTATGCCAAACAAT	ACAAGGAAAAGAAATCAAGAAAGCAGTAC	401
QY	177	GGCGTCTCATCTGGGAAAAAATCTTAAATTTGT	GATGCTTCACAATCTGGAACACATCAA	236
DB	402	GAGCTCTCATCTGGGAAAAGAATCTTAAAGTT	GTGTGCTTCACAACCTGGAGCATTCAA	461
QY	237	TGGGAATGCATTCATATGATCTAGCATGAA	CCATCTCGGAGACATGACTTGTGAAAGAG	296
DB	462	TGGGAATGCATTCATACGATCTGGGCATGA	ACCCTGGGAGCATGACCAGTAGCAAGAG	521
QY	297	TGATATCTTTGATGGCTTCCCTGAGAGTTCC	CAGCAATGGCCACAGAATGCTCACTTATA	356
DB	522	TGATGCTTTTGTATGACTTCCCTGAGAGTTCC	CAGCAATGGCCAGAGAAATATCACATATA	581
QY	357	GGTCAAACTCTAATCAGAAATTGCCTGATTCT	GTGGACTGGAGCAGAGAGGCTGTGTGTA	416
DB	582	AGTCAAAACCCCTAATCCGATATGCCGTGAT	CTCTGGACATGGAGCAGAGAAAGGCTGTGTA	641
QY	417	CTGAAGTGAATACCAAGGTTCTTCTGCTGCT	TGTGTGGCTTTTCAGCGCTGTGGGGGCC	476
DB	642	CTGAAGTGAATATCAAGGTTCTTGTGGTCT	TGCTGGGCTTTTCAGTGTGTGGGGGCC	701
QY	477	TGAAGCACAACCTGAACCTAAAAACAGGA	AGCGTGGTCTCTGAGTGCACAGAACTTGG	536
DB	702	TGAAGCACAACCTGAAGCTGAAACAGGA	AAAGCTGGTGCTCTCAGTGCACAGAACTGG	761
QY	537	TAGATTGCTCAACTGAAAAATATGGGAAT	TAAGGGCTGCAATGGCGGTTTCATGCACACTG	596
DB	762	TGGATTGCTCAACTGAAAAATATGGAAC	AAAGGCCTGCAATGGTGGCTTCATGCACACGG	821
QY	597	CTTTCCAGTATATTATTCATAACACGG	CAATGATTTCAGAGGCTTCTCTATCCCTACAAAG	656
DB	822	CTTTCCAGTATCATTTGATTAACAGGG	CATCGACTCAGAGCGCTTCTCTATCCCTACAAAG	881
QY	657	CCATCAATTGGAACTGCGAGATATGACTT	CAAAAAGCGAGCTGCCCATGTTTCAAGTATA	716
DB	882	CCATGGATCAGAAAATGTCAAATGACT	CAAAAATATCGTGTGCCACATGTTCAAGATACA	941
QY	717	CTGAACCTTCCCTTTGGCAGTGAAGAT	GCCTTAAAAGAGGCTGTGGCCAAATAAAGGACCTG	776
DB	942	CTGAACCTTCCCTTTATGGCAGACAG	AGATGTCCTTAAAAGAGGCTGTGGCCAAATAAAGGACCTG	1001
QY	777	TGCTCTGTGGCTATAGATCGGACCC	ACTATTCTTTCTTCTCTGTACAGAAAGTGGTGTCTACT	836
DB	1002	TGCTCTGTTGGTGTAGATGGGCT	ATCCCTCTTCTTCTCTCTACAGAAAGTGGTGTCTACT	1061
QY	837	ATCAACCACTCTTACTCAGAAATGT	CAATCATGGAGTATTAGTGGTTGGCTATGGTAAACC	896

```

RESULT 3
US-09-652-123-9177
; Sequence 9177, Application US/09652123
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1186-001
; CURRENT APPLICATION NUMBER: US/09/652.123
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/151,135
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 9796
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9177
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2205)
; OTHER INFORMATION: n = A,T,C or G
US-09-652-123-9177

```

Query Match	76.1%;	Score	856.4;	DB	25;	Length	2205;
Best Local Similarity	87.3%;	Fred. No.	1.8e-226;				
Matches	986;	Conservative	0;	Mismatches	136;	Indels	8;
							Gaps
Qy	1	GCATTATTAGTATGGGAGCACCTGGCAGGTCTTTTC	---	ATGAAATGGCTAGTTGGCTGC	57		
Db	222	GGACTCTTACTTGGGAGCAACHTGCTGGTCTATCACAAATGAAACGCTGGTTGTGTGC	281				
Qy	58	TTTCGTTGTGCTCTATGTCAGTGGCACAAGTATCAAGATCCCACTCTCGATCATCACT	117				
Db	282	TCITGGTCTCTCTCTGCACTGGCAGATTGCAATAAAGATCTCTACCTCGATCACCACT	341				
Qy	118	GGAACTCTCTGGAGAAAACCTACAGCAAAACAAT	--CAAGGAAGAGAAATGAGGAAGTACGAC	176			
Db	342	GGCATCTCTGGAGAAAACCTATGGCAACAACAAATACAGGAAAAGAAATGAGAGACGATAC	401				
Qy	177	GGCGTCATCTGGGAAAAAATCTAAAAATTTGTGATGCTTCACAAATCTGCAACACACAA	236				
Db	402	GACGTCTCATCTGGGAAAAGAAATCTTAAAGTTTGTGATGCTTCACAACTGGAGCATTCAA	461				
Qy	237	TGGGAATGCATTTCATGATCTAGGCATGAAACCATCTGGGAGACATGACTGGTGAAGAG	296				
Db	462	TGGGAATGCATCATACATCTGGGCAITGAACCACTGGGAGACATGACCACTGAAGAAG	521				
Qy	297	TCATATCTTTGATGGTTTCCCTGAGAGTTCGCCAGCAATGCCAGAGAAATCTCACTTATA	356				
Db	522	TGATGCTTTGATGAGTTTCCCTGAGAGTTCCAGCCAGTGGCAGAGAAATATCATATATA	581				
Qy	357	GGTCAACTCTTAATCAGAAAATTCCTGATCTTCGACTGGGAGACAGAGGCTGTGTTA	416				

Tue Jul 1 14:40:55 2003

1017 ACCGAGAAATCTA---GACCTCTTCAATTTTATAAAGTCCAAAAATTTGAACACTTTC 1073
 1242 ACCGAGAAATCTAGAGGATCTCTCTTTTATAAAGTCCAAAAATTTGAACACTTTC 1301
 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATTTGTGTCATGA 1122
 1302 TCTTAATTTAATTTTCCGCTGCTATCCAGAGAAATAAATTTGTGTCATGA 1351

RESULT 5
 US-09-710-281-4558
 ; Sequence 4558, Application US/09710281
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John J.
 ; APPLICANT: Stbdal, Hilde
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600.2036-001
 ; CURRENT FILING DATE: 2000-11-10
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 5803
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4558
 ; LENGTH: 2205
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(2205)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-710-281-4558

Query Match 76.1%; Score 856.4; DB 28; Length 2205;
 Best Local Similarity 87.3%; Pred. No. 1.8e-226;
 Matches 986; Conservative

US-09-716-972-4397
 ; Sequence 4397, Application US/09716972
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John J.
 ; APPLICANT: Shyjan, Andrew W
 ; APPLICANT: Vasicek, Thomas
 ; APPLICANT: Lee, John
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600.2030-001
 ; CURRENT FILING DATE: 2000-11-21
 ; PRIOR FILING DATE: 1999-11-22
 ; NUMBER OF SEQ ID NOS: 5251
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4397
 ; LENGTH: 2205
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(2205)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-716-972-4397

Query Match 76.1%; Score 856.4; DB 28; Length 2205;
 Best Local Similarity 87.3%; Pred. No. 1.8e-226;

1017 ACCGAGAAATCTA---GACCTCTTCAATTTTATAAAGTCCAAAAATTTGAACACTTTC 1073
 1242 ACCGAGAAATCTAGAGGATCTCTCTTTTATAAAGTCCAAAAATTTGAACACTTTC 1301
 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATTTGTGTCATGA 1122
 1302 TCTTAATTTAATTTTCCGCTGCTATCCAGAGAAATAAATTTGTGTCATGA 1351

RESULT 6
 US-09-716-972-4397
 ; Sequence 4397, Application US/09716972
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John J.
 ; APPLICANT: Shyjan, Andrew W
 ; APPLICANT: Vasicek, Thomas
 ; APPLICANT: Lee, John
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600.2030-001
 ; CURRENT FILING DATE: 2000-11-21
 ; PRIOR FILING DATE: 1999-11-22
 ; NUMBER OF SEQ ID NOS: 5251
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4397
 ; LENGTH: 2205
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(2205)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-716-972-4397

Query Match 76.1%; Score 856.4; DB 28; Length 2205;
 Best Local Similarity 87.3%; Pred. No. 1.8e-226;
 Matches 986; Conservative

1017 ACCGAGAAATCTA---GACCTCTTCAATTTTATAAAGTCCAAAAATTTGAACACTTTC 1073
 1242 ACCGAGAAATCTAGAGGATCTCTCTTTTATAAAGTCCAAAAATTTGAACACTTTC 1301
 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATTTGTGTCATGA 1122
 1302 TCTTAATTTAATTTTCCGCTGCTATCCAGAGAAATAAATTTGTGTCATGA 1351

RESULT 5
 US-09-710-281-4558
 ; Sequence 4558, Application US/09710281
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John J.
 ; APPLICANT: Stbdal, Hilde
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600.2036-001
 ; CURRENT FILING DATE: 2000-11-10
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 5803
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4558
 ; LENGTH: 2205
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(2205)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-710-281-4558

Query Match 76.1%; Score 856.4; DB 28; Length 2205;
 Best Local Similarity 87.3%; Pred. No. 1.8e-226;
 Matches 986; Conservative

1017 ACCGAGAAATCTA---GACCTCTTCAATTTTATAAAGTCCAAAAATTTGAACACTTTC 1073
 1242 ACCGAGAAATCTAGAGGATCTCTCTTTTATAAAGTCCAAAAATTTGAACACTTTC 1301
 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATTTGTGTCATGA 1122
 1302 TCTTAATTTAATTTTCCGCTGCTATCCAGAGAAATAAATTTGTGTCATGA 1351

RESULT 6
 US-09-716-972-4397
 ; Sequence 4397, Application US/09716972
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter, John J.
 ; APPLICANT: Shyjan, Andrew W
 ; APPLICANT: Vasicek, Thomas
 ; APPLICANT: Lee, John
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; FILE REFERENCE: 1600.2030-001
 ; CURRENT FILING DATE: 2000-11-21
 ; PRIOR FILING DATE: 1999-11-22
 ; NUMBER OF SEQ ID NOS: 5251
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4397
 ; LENGTH: 2205
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(2205)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-716-972-4397

Query Match 76.1%; Score 856.4; DB 28; Length 2205;
 Best Local Similarity 87.3%; Pred. No. 1.8e-226;
 Matches 986; Conservative

Matches	986:	Conservative	0:	Mismatches	136:	Indels	8:	Gaps	4:
Qy	1	GCATTATTACTATGCGGACCACTGCAGGTGCTTTTC---	ATCAAAATGGCTAGTTGGGTGC	57					
Db	222	GGACTCTTACTGTGGGACCACTGCTGGTTCTATCAAAATCAAAACGGCTGGTTGTGTCGC	281						
Qy	58	TTCCGTTTGCTCTCTATGCGAGTGGCACAAGTACATAAAGATCCCACTCTGGATCATCACT	117						
Db	282	TCATTGGTGTGCTCTCTGCGAGTGGCACAAGTGGCAATAAAGATCCCTACCACTCAGTCCACT	341						
Qy	118	GGAAATCTCTGGAGAAAACCTTACAGCAAAACAAT--CAAGGAAGACAAATCGAGAACTAGCAC	176						
Db	342	GGCATCTCTGGAGAAAACCTATGGCAACAATACAGAGAAAAGATGAAGAAGCACTGAC	401						
Qy	177	GGCGTCTCATCTGGGAAAAAATCTAAAAATTTGATGCTTTCACAAATCTGGAACACTCAA	236						
Db	402	GACCTCTCATCTGGGAAAAGAATCTAAAGTTTGTGATGCTTCACAAGCTGGAGCATTCAA	461						
Qy	237	TGGCAATGCAATTCATATCATCTAGCGATGAACCATCTGGAGATATCATCTGGTAAAGAAG	296						
Db	462	TGGGAATGCCATCATACGATCTGGGCATGAACCACTTGGAGACATGACCAGTGAAGAAG	521						
Qy	297	TGATATCTTTGATGGGTTTCCTGAGAGTTCCCAAGCCAAATGGCAGAAAATGTCTCATTTATA	356						
Db	522	TGATGTCTTTGATGAGTTCGCTGAGAGTTCCCAACAGTGGCAGAGAAAATATCACATATA	581						
Qy	357	GGTCAAAACTCTTAATTCAGAAATTCGCTCATCTGTGGACTGGAGACAGAAGGCGTCTGTA	416						
Db	582	AGTCAAAACCCCAATGCGGATATTCGCTCATCTGTGGACTGGAGACAGAAGGCGTCTGTA	641						
Qy	417	CTGAAGTGAATACACAGGGTTCTTTGGTGTCTTTGGGCTTTTCAGCGCTGTGGGGGCC	476						
Db	642	CTGAAGTGAATATCAAGGTTCTTTGGTGTCTTTGGGCTTTTCAGTCTCTGGGGGCC	701						
Qy	477	TGGAAGCACAACTGAAGCTAAAAACAGGAAGCTGGTCTCTCAGTGCACAGAACTTGG	536						
Db	702	TGGAGCACAGCTGAAGCTGAAAACAGGAAGCTGGTCTCTCAGTGCACAGAACTTGG	761						
Qy	537	TGATTTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGGTTTCATGACAACTG	596						
Db	762	TGGATTTGCTCAACTGAAAAATATGGGAACAAAGGCTGCAATGGTGGCTTCATGACAACTG	821						
Qy	597	CTTTCCAGTATATTATTCAT	656						
Db	822	CTTTCCAGTAT	881						
Qy	657	CCATGAATGGAAGCTGCAGATATGACATCAAAAAAGCGAGCTGGCACATGTTTCAAAAGTATA	716						
Db	882	CCATGGATCAGAAATGTCATATGACTCAAAATATATATATATATATATATATATATATAT	941						
Qy	717	CTGAACHTTCCGTTTGGCAGTCAAGATGCCCTTAAAGAAAGCTGTGGCCCAATAAAGGACCTG	776						
Db	942	CTGAACHTTCCCTATTGGCAGAGACATGCTCTGAAAGAAAGCTGTGGCCCAATAAAGGACCTG	1001						
Qy	777	TGCTTTGCTATAGATGCGAGGCACTATTCTTTCTCTGTACAGAAAGTGGTGTCTACT	836						
Db	1002	TGCTGT	1061						
Qy	837	ATCAACCATCTGTACTCAGAATGTGAATCATGCAATATAGTGGTTGGCTATGGTAAC	896						
Db	1062	ATGAACCATCTGTACTCAGAATGTGAATCATGCAATATAGTGGTTGGCTATGGTAAC	1121						
Qy	897	TTAATCGGAAGACTACTGGCTTGTGAAAACAGCTGGGGCCTCAACTTTGGTGAACCAAG	956						
Db	1122	TTAATCGGAAGAAATACTGGCTTGTGAAAACAGCTGGGGCCTCAACTTTGGTGAACCAAG	1181						
Qy	957	GATATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1016						
Db	1182	GATATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1241						
Qy	1017	ACCAGAAATCTA---GACCTCTTCATTTTATTAACAAGTCCAAAAATTAAGAACACTTTC	1073						
Db	1242	ACCAGAAATCTAGAGATCTCTCTTTTATTAACAATTAAGAACACTTTCATTAAGAACACTTTC	1301						

```

Qy 1074 TCTTAATTTAAATTTTACTGCTGATA-ATAGTAAAAAATAATGTGTCATGA 1122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1302 TCTTAAGCTTAATTTTCTGCTGTATCCAGAAGAAATAAATTCGTGTCATGA 1351
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-716-990-1545
; Sequence 1545, Application US/09716990
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2039-001
; CURRENT APPLICATION NUMBER: US/09/716,990
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/166,926
; PRIOR FILING DATE: 1999-11-22
; NUMBER OF SEQ ID NOS: 1691
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1545
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2205)
; OTHER INFORMATION: n = A,T,C or G
US-09-716-990-1545

```

Query Match	Score	DB 28:	Length
Best Local Similarity	76.1%	856.4	2205
Matches 986: Conservative	87.3%	Pred. NO. 1.80-226:	
Matches 986: Conservative	0:	Mismatches 136:	Gaps 4:

Db	762	TGGATTTGCTCAACTGAAATAATGAAAAAAGAGGTGCAATGGTGGCTTCATGACAAACGG	821
QY	597	CTTTCACGATATATTATGATAACACAGGCAATTGATTACAGAGCTTTCCTATCCCTACAAAG	656
Db	822	CTTTCACGATCATATTGATACACAGGCAATGACTCAGACGTTCCCTATCCCTACAAAG	881
QY	657	CCATGAATGGAAAGTGCAGATATGACTCAAAAAGCGAGCTGCCACATGTTCAAAAGTATA	716
Db	882	CCATGGATCAAAATGTCNAATATGATCAAAATATCGTCTGCCACATGTTCCAAAGTACA	941
QY	717	CTGAACCTCCCTTTGGCAGTGAAGATGCCCTTAAAAAGAGCTGTGCCCAATAAAGGACCTG	776
Db	942	CHCAACTTCCTTATGGCAGAGAAGATGCTCGAAGAACGCTGTGCCCAATAAAGGCCAG	1001
QY	777	TGTCGTGGCTATAGATGGAGCGACCTATCTTTCTTCCTGTACAGAGTGGTGCTACT	836
Db	1002	TGTCGTGGTGTAGATGGCGGTCACTCTCTCTCTCTGTACAGAGTGGTGCTACT	1061
QY	837	ATGAACCATCTCTACTCAGAAATGTAATCATCGAGTATTAGTGGTTGGCTATGGTAAACC	896
Db	1062	ATGAACCATCTCTACTCAGAAATGTAATCATCGAGTACTTGTGGTTGGCTATGGTGATC	1121
QY	897	TTAATGGGAAGACTACTGGCTTGTGAAAACACAGCTGGGGCCTCAACTTTGGTGACCAAG	956
Db	1122	TTAATGGGAAGAAATACTGGCTTGTGAAAACACAGCTGGGGCCTCAACTTTGGTGAAAGAAG	1181
QY	957	GATATATTCGGATGGCAAGAACAGTGGAAATCACTGTGGGATTGCTAGTTATCCCTCTT	1016
Db	1182	GATATATTCGGATGGCAAGAAATCACTGTGGGATTGCTAGTTATCCCTCTT	1241
QY	1017	ACCAGAAATCTA---GACCTCTTCAATTTTATTAACAGTCCCAAAAATTTGAAACACTTTC	1073
Db	1242	ACCAGAAATCTAAGAGATCTCTCTCTTTTATTAACAAATCAGAAATATGAAGCACATTC	1301
QY	1074	TCCTAATTAATTTTACCTGCTATA-ATAGTAAAAATAAAATGTGTCATGA	1122
Db	1302	TCCTAATTAATTTTCCGCTGATCCAGAACAAATAATTTGTCATGA	1351
RESULT 8			
US-09-721-589-5798			
: Sequence 5798, Application US/09721589			
: GENERAL INFORMATION:			
: APPLICANT: Gearing, David P.			
: APPLICANT: Villevial, Jean-Luc			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
: TITLE OF INVENTION: THEREFOR			
: FILE REFERENCE: 1600.2045-001			
: CURRENT APPLICATION NUMBER: US/09/721.589			
: CURRENT FILING DATE: 2000-11-22			
: PRIOR APPLICATION NUMBER: 60/167,380			
: PRIOR FILING DATE: 1999-11-24			
: NUMBER OF SEQ ID NOS: 7017			
: SOFTWARE: FastSeq for Windows Version 4.0			
: SEQ ID NO 5798			
: LENGTH: 2205			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (1)..(2205)			
: OTHER INFORMATION: n = A,T,C or G			
US-09-721-589-5798			

RESULT 9

US-09-726-172-2249

```

: Sequence 2249, Application US/09/726172
:
: GENERAL INFORMATION:
:
: APPLICANT: Gearing, David P.
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2052-001
: CURRENT APPLICATION NUMBER: US/09/726.172
: CURRENT FILING DATE: 2000-11-29
: PRIOR APPLICATION NUMBER: 60/167.846
: PRIOR FILING DATE: 1999-11-29
: NUMBER OF SEQ ID NOS: 2733
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 2249
: LENGTH: 2205
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1) (2205)
: OTHER INFORMATION: n = A,T,C or G
:
: US-09-726-172-2249

```

Query Match 76.1%; Score 856.4; DH 29; Length 2205;
Best Local Similarity 87.3%; Pred. NO. 1.8e-226;
Matches 986; Conservative 0; Mismatches 136; Indels 8; Gaps 4;

QY	1	GCATATTAGTATGGGAGACGCTCGAGGTTCTTTCC---ATGAAATGGCTAGTTGGGCTGC	57
DB	222	GGACCTCTTACTGTGGGAGCAACTCGTGGTTCTATCACAATGAACAGGCTGGTTTGTGTGC	281
QY	58	TTCCGTTTGTGCTCTATGCACTGGCACTCAAGTACATAAAGATCCACTCTCGGATCATCACT	117
DB	282	TCTTGGTGTGCTCTCTGCACTGGCACAGTTGGATAAAGATCTCCTTGGATCACCCT	341
QY	118	GGATCTCTGGAAGAAACCTTACAGCAACAAT-CAAGCAAGAGAAATAGGAGATAGCAC	176
DB	342	GGCATCTCTGGAAGAAACCTTATGGCAACAATACAAAGGAAAAAGAAATGAAGAAGCAGTAC	401
QY	177	GGCGTCTCATCTGCAAAAAAATCTCAAAATTTGTGATGCTTCAATCTGGGAACACTCAA	236
DB	402	GACGCTCTCATCTGGGAAAAAGAACTAAAGTTTGTGATGCTTCAACACCTGGAGCATTCAA	461
QY	237	TGGGAATGCAATCATATGATATAGGATGAAACATCTGGGAGACATGATGGTGAAGAG	296
DB	462	TGGGAATGCACCTCATAGCATCTGGGCATGAACACCTGGGAGACATGACGATGAAGAAG	521
QY	297	TGATATCTTTCATATGCGTTGCGCTCGAGCTTCCGAGCCAAATGGCAGAGAAATCTCACTTATA	356
DB	522	TGATGCTCTTATGAGTTCCCTGAGATTCGCCAGCAGTGGCAGAGAAATATACATATATA	581
QY	357	GGTCAAACTCTAAATCTCAAAATTTGCTGATTTGTGGACTGGGAGAGAAAGGCTGTGTTA	416
DB	582	ACTCAAAACCTTAATCGGATATTGCCCTGATCTCTGGACTGGCAGAGAAAGGCTGTGTTA	641
QY	417	CTGAAGTCAAAATACAGAGATCTTGTAAGTGTTGTTGAGGCTTTCAGAGTGTGGGGGGCC	476
DB	642	CTGAAGTGAATAATCAAGGTTCTTGTAAGTCTGCTGAGGTTTTCAGTGTGTGGGGGGCC	701
QY	477	TGGAAGCACAACTCAAGCTAAAAACAGGAAGCTGCTGCTCTGAGTGGCAGACAGAACTTGG	536
DB	702	TGGAAGCACACTGAAGCTGAAACAGAAAGTGTGTCTCTCATGTGCCAAGAACTGG	761
QY	537	TAGATTGCTCAACTGAAAAATATGGGAATTAAGGCTGCCAATGGCGGTTTCATGACAAC	596
DB	762	TGGAATGCTCAACTGAAAAATATGGAACAAGGCTGCCAATGGTGGCTTCATGACAACGG	821
QY	597	CTTTCCAGTATATTATGATACAAAGGCATTGATTCAGAGAGCTTCTATCCCTACAAG	656
DB	822	CTTTCCAGTATCATTTGATTAACAAAGGGATTCGACTTCAGAGCTTCTCTATCCCTACAAG	881
QY	657	CCATGAATGGAAGCTGCAGATATGACTCAAAAAAGGAGGCTGCCACATGTTCAAAAGTATA	716

us-10-010-580-1.rnrm

Tue Jul 1 14:40:55 2003

Job time : 2968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:28:21 ; Search time 978 Seconds

(without alignments)
8229.219 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1126

Sequence: 1 gcattattagtgaggagca.....ataaatgtgtcatgaccat 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched 8166257 seqs. 3573790582 residues

Total number of hits satisfying chosen parameters: 1633514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA, New: *

- 1: /cgn2_6/ptodata/2/pna/pct_NEW_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match %	Length	ID	Description
1	856	76.0	2325	8	US-10-227-425-1251
2	856	76.0	2350	8	US-10-227-425-25
3	854.8	75.9	1514	10	US-10-170-235-26187
4	853.6	75.8	1514	12	US-60-452-680-10960
5	853.6	75.8	1514	12	US-60-453-135-6918
6	853.6	75.8	1514	12	US-60-453-050-6918
7	853.6	75.8	1514	13	US-60-455-444-3608
8	853.6	75.8	1514	13	US-60-465-241-3608
9	853.6	75.8	1514	13	US-60-466-412-6918
10	845.4	75.1	1763	5	US-09-873-367C-173
11	845.4	75.1	1763	10	US-10-219-051B-2756
12	845.4	75.1	1763	10	US-10-219-051B-10848
13	842.4	74.8	1255	6	US-10-342-887-987
14	824.8	73.3	1366	6	US-09-724-676-33122
15	824.8	73.3	1366	6	US-09-724-676A-33122
16	824.8	73.3	1409	6	US-09-724-676-33121
17	824.8	73.3	1409	6	US-09-724-676A-33121
18	709.2	63.0	1895	1	PCT-US03-09774-96
19	665.8	59.1	1290	9	US-10-144-771-6528
20	634.2	56.3	1347	10	US-10-256-911-157

c

ALIGNMENTS

RESULT 1

US-10-227-425-1251

; Sequence 1251, Application US/10227425

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT249CIN

; CURRENT APPLICATION NUMBER: US/10/227,425

; CURRENT FILING DATE: 2002-08-26

; PRIOR APPLICATION NUMBER: 09/760,475

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: 60/214,886

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/217,487

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,758

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/220,963

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/217,496

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,447

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/218,290

; PRIOR FILING DATE: 2000-07-14

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1251
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-227-425-1251
Query Match 76.0%, Score 856, DB 8, Length 2325;
Best Local Similarity 87.2%, Pred. No. 14e+20,
Matches 985; Conservative 1; Mismatches 136; Indels 8; Gaps 4;

QY

1 GCATTATTAGTATGGGACCACTCGAGTCTTTC---ATCAATGCTAGTTCGGCIGC 57

Sequence 156, App
Sequence 3409, App
Sequence 2945, App
Sequence 1700, App
Sequence 949, App
Sequence 2002, App
Sequence 155, App
Sequence 3409, App
Sequence 2754, App
Sequence 10846, A
Sequence 2945, App
Sequence 831, App
Sequence 8754, App
Sequence 243045, App
Sequence 1, Appl1
Sequence 1556, App
Sequence 479, App
Sequence 6305, App
Sequence 11530, A
Sequence 200, App
Sequence 26186, A
Sequence 10959, A
Sequence 6917, App
Sequence 6917, App
Sequence 3607, App

Db 86 GGACTCTTACTGTRGAGCAACTGCTGGTCTTATCACAATGAACGGCTGTTGTGTC 145
QY 58 TTCCTGTGCTCCTATGAGTGGCACAAGTACATAAAGATCCCACTCTGATCATCACT 117
Db 146 TCTTGGTGTGCTCCTCCTCCTGAGTGGCAGCTGCTGATTAAGATCCCTACCTGATCA 205
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAC 176
Db 206 GGCATCTCTGGAAGAAACCTATGCAACAATAACAGGAAGAAATGAGGAAGTAC 265
QY 177 GCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTCATGCTTCCACAACTGGAACACTCA 236
Db 266 GAGCTCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCCACAACCTGGAGCATCA 325
QY 237 TGGGAATGATTCATATGATCTAGGCATGAACCATCTGGAGACATGACTGGTGAAGAG 296
Db 326 TGGGAATGACTCATACGATATTCGCTGATTCGCTGAGTGGGAGACATGACTGAGGAAG 385
QY 297 TCATATCTTTGATGGTTCCTCCAGAGTTCCAGGCAATGGCAGAGAAATGTCACTTATA 356
Db 386 TGATGCTTTGATGAGTTCCCTGAGATTCAGGCTGAGGAGGAGAAATATCACATATA 445
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGTGCTGAGTGGAGAGAGGCTGTGTTA 416
Db 446 ACTCAACCTAATCGGATATTCGCTGATTCGCTGAGTGGGAGAGAGGAGGCTGTGTTA 505
QY 417 CTGAAGTGAATACCAAGGTTCTGTTGGTGTGTTGGCTTTCAGCGCTTGGGGGCC 476
Db 506 CTGAAGTGAATATCAAGGTTCTGTTGGTGTGTTGGCTTTCAGCGCTTGGGGGCC 565
QY 477 TGAAGCACTGAAGCTAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 536
Db 566 TGAAGCAGAGTGAAGCTGAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 625
QY 537 TAGATTGCTCAACTGAAATATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 596
Db 626 TGATTTGCTCAACTGAAATATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 685
QY 597 CTTTCAGTATATTATGATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 656
Db 686 CTTTCAGTATATTATGATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 745
QY 657 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAG 716
Db 746 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAG 805
QY 717 CTCACCTTCCCTTTGGCAGTGAAGATGCTTAAAAAGAGGAGGAGGAGGAGGAGGAGG 776
Db 806 CTGAACCTTCCCTTTGGCAGTGAAGATGCTTAAAAAGAGGAGGAGGAGGAGGAGGAG 865
QY 777 TGTCTGTGCTATAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 836
Db 866 TGTCTGTGCTATAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 925
QY 837 ATGAACCTTCCCTTTGGCAGTGAAGATGCTTAAAAAGAGGAGGAGGAGGAGGAGGAG 896
Db 926 ATGAACCTTCCCTTTGGCAGTGAAGATGCTTAAAAAGAGGAGGAGGAGGAGGAGGAG 985
QY 897 TTAATGGGAAGAGTACTGGCTTGTCAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 956
Db 986 TTAATGGGAAGAGTACTGGCTTGTCAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1045
QY 957 GATATATTGGATGGCAAGAAACAGTGGAAATCATCTGAGGAGTGGTATGCTATGCTCT 1016
Db 1046 GATATATTGGATGGCAAGAAACAGTGGAAATCATCTGAGGAGTGGTATGCTATGCTCT 1105
QY 1017 ACCCAAAATCTA---GACCTTTCATTTTAAACAGTCCAAAAGTAAAGAAATGAGGAG 1073
Db 1106 ACCCAAAATCTA---GACCTTTCATTTTAAACAGTCCAAAAGTAAAGAAATGAGGAG 1165
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 1122
||||| ||||||| ||||||| || ||||||| |||||||

Db 1166 TCTTAACCTAATTTTCTCTCTGTATCCAGAGAAATAATTTGTCATGA 1215
RESULT 2
US-10-227-425-25
; Sequence 25, Application US/10227425
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PIZ49CIN
; CURRENT APPLICATION NUMBER: US/10227425
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/760,475
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-227-425-25

Query Match 76.08; Score 856; DH 8; Length 2350;
Best Local Similarity 87.28; Pred. No. 1.4e-220;
Matches 985; Conservative 1; Mismatches 116; Indels 8; Gaps 4;
QY 1 GCATTAT 57
Db 111 GGACTCTTACTGTGGAGCAACTGCTGCTTCTATCAATGAACGGCTGTTGTGTC 170
QY 58 TTCCGTTGCTGCTCTATGCTGAGTGGCACAAGTACATAAAGATCCCACTCTGATCATCACT 117
Db 171 TCTTGGTGTGCTCTATGCTGAGTGGCACAAGTACATAAAGATCCCACTCTGATCATCACT 240
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAC 176
Db 231 GGCATCTCTGGAAGAAACCTATGCAACAATAACAGGAAGAAATGAGGAAGTAC 240
QY 177 GCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTCATGCTTCCACAACTGGAACACTCA 236
Db 291 GAGCTCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCCACAACCTGGAGCATCA 350
QY 237 TGGGAATGATTCATATGATCTAGGCATGAACCATCTGGAGACATGACTGGTGAAGAG 296
Db 351 TGAATATGCA-TATATGAT-TGAG-AT-AAA-AT-TGAGAA-ATGAGGAGTGAAGAG 410
QY 297 TGATATCTTTGATGGTTCCTGAGAGTTCGCCAGCAATGGCAGAGAAATGTCATTTATA 356
Db 411 TGATCTCTTTGATGGTTCCTGAGAGTTCGCCAGCAATGGCAGAGAAATGTCATTTATA 470
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGTGAGTGGAGAGGAGGAGGAGGAGGAG 416
Db 471 AGTCAAACTCTAATCAGGATATTCCTGATCTGTGAGTGGAGAGGAGGAGGAGGAGGAG 530

QY	417	CTGAAGTGAATACCAAGGTTCTCTGTGGTCTGTGTGGCTTTACGCGCTGTGGGGCCC	476
DB	531	CTCAAGTGAATATCAAGTTCCTGTGTGTCTGTCTGGCTTTTCAGTGTGTGGGGCCC	590
QY	477	TGGAACGACACTGAAGCTTAAACACAGGAAGCTGTGTCTGTAGTGCACAGAACTGG	536
DB	591	TGGAAGCAGACTGGAAGCTTGAACACAGGAAGCTGGTGTCTCTAGTGCACAGAACTGG	650
QY	537	TGATTTGCTCAACTGAAAAATATGGAAATTAAGGCTGCAATGGCGGTTTCATGCACAAC	596
DB	651	TGCAATTCCTCAACTGAAAAATATGGAACAAAGGCTGCAATGCTGGCTTTCATGCACAAC	710
QY	597	CTTTCAGTATATTATGATAACAACGGCATTGATTCAGAAGTTCCTTATCCCTCAACAAG	656
DB	711	CTTTCAGTATACATTTGATAACAGGGATGCACCTACAGCGTTCCTATCCCTCAACAAG	770
QY	657	CCATGAATGAAAGTGCACATATGACTTCAAAAAGGAGCTGCACATGTTCAAGATATA	716
DB	771	CCATGATCAGAAATGTCAATATGACTCAAAATATCTGTGTGCACATGTTCAAGATACA	830
QY	717	CTGAACCTCCCTTTGGCAGTGAAGATGCCCTTAAAGAAAGCTGTGGCCATTAAGGAGCTG	776
DB	831	CTGAACCTCCCTTTATGGCAGAGAAGATGTCTGAAAGAAAGCTGTGGCCATTAAGGAGCTG	890
QY	777	TGCTGTGGCTATAGATGCGAGCCACTATTCTTCTCTGTACAGAAAGTGGTGTCTACT	836
DB	891	TGCTGTGTGTGTAGATCCCGGTACCTCTTCTTCTCTACAGAAAGTGGTGTCTACT	950
QY	837	ATGAACCATCCTGTACTCAGAATGTGAATCATGGAGTATTTAGTGTGTGGCTATGGTAACC	896
DB	951	ATGAACCATCCTGTACTCAGAAATGTGAATCATGCTGTACTGTGTGTGGCTATGGTGATC	1010
QY	897	TTAATGGAAGACTACTGGCTTGTGAAACACAGCTGGGCGCTCAACTTTGGTGCACCAAG	956
DB	1011	TTAATGGAAGAAATACTGGCTTGTGAAACACAGCTGGGCGCCACAACTTTGGTGAAGAAG	1070
QY	957	GATATATTGGATGGCAAGAACAGTGGAAATCACTGTGGATTGTAGTTATCCCTCTT	1016
DB	1071	GATATATTGGATGGCAAGAAATTAAGGAAATCATTTGTGGGATTGTCTAGCTTTCCCTCTT	1130
QY	1017	ACCCAGAAATCTA--GACCTCTTCATTTTATAACAAGTCCAAAAATTTGAACACTTTC	1073
DB	1131	ACCCAGAAATCTAGAGGATCTCTCTCTTTTATACAATCAAGAAATATGAGCACATTC	1190
QY	1074	TCCTTAATTTAAATTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA	1122
DB	1191	TCCTTAACCTAAATTTTCTGCTGTATCCAGAAGAAATTAATTTGTGTCATGA	1240

RESULT 3

RES/DT 3
115-10-170-235-26187US-10-170-233-26187
: Commerce 26187 : A; Sequence 26187, Appl
GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig

; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF

;; TITLE OF INVENTION: TRANSCRIPTS, FOR DET

FILE REFERENCE: CL001380

: CURRENT APPLICATION NUMBER: US/1

CURRENT AFFILIATION NUMBER: 2003-
CURRENT FILING DATE: 2003-

NUMBER OF SEQ

; NUMBER OF SEQ
SEQ ID NO 76197

; SEQ ID NO 26187

```
; LENGTH: 1514
```

; TYPE: DNA

ORGANISM: HUMAN

US-10-170-235-26187

3
3
1
2
2
4
2
4
1
5
6

```
Query Match
75.98; Score 854.8; DB 10; Length 1514;
```

Query Match 75.9%; Score 0.54.0; BB 10; Length 1017;
Best Local Similarity 87.2%; Pred. No. 2.6e-220;

```

RESULT 4
US-60-452-680-10960
; Sequence 10960, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GROUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10960
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-10960

Query Match      75.8%; Score 853.6; DB 12; Length 1514;
Best Local Similarity 86.6%; Pred No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACCTGGAGTTCTTTC---ATGAATGGCTAGTTGGGCTGC 57
DB 224 GGACTCTTACTCTGGAGCACTGGCTGTCTATCACAATGAACGGCTGGTTGTGCGC 283
QY 58 TTCCGTTGTCTCTATGCGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 284 TCTTGGTGTCTCTGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 343
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGC 176
DB 344 GGCACTCTGGAAGAAACCTATGGCAACAATCAAGGAAGAAATGAGGAAGTAGC 403
QY 177 GFGCTCTATCTGGGAAAGAAATCTAATAATTTGTGATGCTTCACATCTGGACACTCA 236
DB 404 GAGCTCTATCTGGGAAAGAAATCTAATAATTTGTGATGCTTCACACCTGGAGCACTCA 463
QY 237 TGGGAATGCAATCATATGATCTAGGCATGAACCACTCTGGGAGACATGATGGTGAAGA 296
DB 464 TGGGAATGCAATCATATGATCTAGGCATGAACCACTCTGGGAGACATGATGGTGAAGA 523
QY 297 TGAATCTCTTGTGGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAGAGT 356
DB 524 TGATGCTCTTGTGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAGAGT 583
QY 357 GGTCAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTGA 416
DB 584 AGTCAAAACCTAATGAGATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTGA 643
QY 417 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAG 476
DB 644 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAG 703
QY 477 TGAAGCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 536
DB 704 TGAAGCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 763
QY 537 TAGATGCTCAACTGAGAAATATGGAATTAAGGCTGCAATGGCGGTTTCATGACAAGT 596
DB 764 TGAATGCTCAACTGAGAAATATGGAATTAAGGCTGCAATGGCGGTTTCATGACAAGT 823
QY 597 CTTTCAGTATATATTGATACACAGGCAATGATTCAGAACCTTCCTATCCCTACAAG 656
DB 824 CTTTCAGTATATATTGATACACAGGCAATGATTCAGAACCTTCCTATCCCTACAAG 883
QY 657 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 716
DB 884 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 943
QY 717 CTGAACCTTCCCTTTGGCACTGAAGATGGCTTTAAAAGAAAGCTGTGTGCCCAATAAGAG 776

```

```

DB 944 CTGAACCTTCCCTTTATGSSAGAAAGATGCTGGAAGAGCTGTGGCAATAAAGAGCCAG 1004
QY 777 TGTCTGTGGCTATAGATCCGAGCCACTATTTCTTCTCTGCTACAGAGTGTGTCTACT 846
DB 1004 TGTCTGTGGCTATAGATCCGAGCCACTATTTCTTCTCTGCTACAGAGTGTGTCTACT 1064
QY 837 ATGACCACTCTCTACTTACTAGATGTGAATATGAGAGTATTAAGTGTGTGTACTAAG 896
DB 1064 ATGACCACTCTCTACTTACTAGATGTGAATATGAGAGTATTAAGTGTGTGTACTAAG 1124
QY 897 TTAATGGGAAAGACTTACTGCTTGTGAAAGAAAGCTGGAGGCTCAATTTTGGTGAAG 956
DB 1124 TTAATGGGAAAGAAATATGGCTTGTGAAAGAAAGCTGGAGGCTCAATTTTGGTGAAG 1184
QY 957 CATATATTCGGATGGCAAGAACAGTGGAAATCACTGGGAGTGTGAGTATGCTCTCT 1016
DB 1184 CATATATTCGGATGGCAAGAAATTAAGGAAATCAATTTGGGATTTGTAGTTTCTCT 1244
QY 1017 ACCAGAAATCTA---GACCTCTTCATTTTATAACAGTGTGMAAAATTTGAAATCTT 1074
DB 1244 ACCAGAAATCTAGAGATCTCTCTTTTATAACAATCAAGAAATATCAAGGACTTCT 1404
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATTAATTTGTGTCATGA 1122
DB 1304 TCTTAATTTAATTTTCTGCTGCTATGCAAGAAATAATTTGTGTCATGA 1454

RESULT 5
US-60-453-135-6918
; Sequence 6918, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: JAKOUHOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6918
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-453-135-6918

Query Match      75.8%; Score 853.6; DB 12; Length 1514;
Best Local Similarity 86.6%; Pred No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACCTGGAGTTCTTTC---ATGAATGGCTAGTTGGGCTGC 57
DB 224 GGACTCTTACTCTGGAGCACTGGCTGTCTATCACAATGAACGGCTGGTTGTGCGC 283
QY 58 TTCCGTTGTCTCTATGCGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 284 TCTTGGTGTCTCTGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 343
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGC 176
DB 344 GGCACTCTGGAAGAAACCTATGGCAACAATCAAGGAAGAAATGAGGAAGTAGC 403
QY 177 GFGCTCTATCTGGGAAAGAAATCTAATAATTTGTGATGCTTCACATCTGGACACTCA 236
DB 404 GAGCTCTATCTGGGAAAGAAATCTAATAATTTGTGATGCTTCACACCTGGAGCACTCA 463
QY 237 TGGGAATGCAATCATATGATCTAGGCATGAACCACTCTGGGAGACATGATGGTGAAGA 296
DB 464 TGGGAATGCAATCATATGATCTAGGCATGAACCACTCTGGGAGACATGATGGTGAAGA 523
QY 297 TGAATCTCTTGTGGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAGAGT 356
DB 524 TGATGCTCTTGTGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAGAGT 583
QY 357 GGTCAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTGA 416
DB 584 AGTCAAAACCTAATGAGATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTGA 643
QY 417 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAG 476
DB 644 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGTTCCTCAGAGTTCCTCAG 703
QY 477 TGAAGCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 536
DB 704 TGAAGCAACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 763
QY 537 TAGATGCTCAACTGAGAAATATGGAATTAAGGCTGCAATGGCGGTTTCATGACAAGT 596
DB 764 TGAATGCTCAACTGAGAAATATGGAATTAAGGCTGCAATGGCGGTTTCATGACAAGT 823
QY 597 CTTTCAGTATATATTGATACACAGGCAATGATTCAGAACCTTCCTATCCCTACAAG 656
DB 824 CTTTCAGTATATATTGATACACAGGCAATGATTCAGAACCTTCCTATCCCTACAAG 883
QY 657 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 716
DB 884 CCATGAATGGAAGTCCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 943
QY 717 CTGAACCTTCCCTTTGGCACTGAAGATGGCTTTAAAAGAAAGCTGTGTGCCCAATAAGAG 776

```



```
QY 1017 ACCGAGAAATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAATTTGAACACACTTTC 1073
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1244 ACCGAAATCTAGAGGATCTCTCTCTTTTATAACAATCAAGAAATATGAGGACACTTTC 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTGCATGA 1122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1304 TCTTAATTTAATTTTCTGCTGTATCCAGAGAATAAATTTGTGTGCATGA 1353

RESULT 7
US-60-455-444-3608
; Sequence 3608, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001455
; CURRENT APPLICATION NUMBER: US/60/455,444
; CURRENT FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3608
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-455-444-3608

Query Match 75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGACGACCTCGCAGGTTCTTTC---ATGAATTTGGCTAGTTGGGCTGC 57
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 GCACCTTACTCTRGAGCAACTGCTGCTCTATCACAATGAAGCGGCTGGTTGTGCGC 283

QY 58 TTCGTTGTGCTCTATGCAATGGGACAAAGTACATAAAGATCCCTCTGTGATCATCAT 117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 343

QY 118 GGAATCTCTGGAAGAAACCTTACAGCAACAAT-CAAGGAAGAGAAATCAGGAAGTAGCAC 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 344 GCATCTCTGGAAGAAACCTTATGCAACAATACAGGAAGAAATGAGGACGATAC 403

QY 177 GCGTCTCATCTGGGAAAAAATCTAAATTTTGTGATGCTTCCACAACTCGAAGCACTCAA 236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 GACGTCTCATCTCGGAAAAAATCTAAAGTTTGTGATGCTTCCACAACTCGGAGCATTCAA 463

QY 237 TGGGAATGCATTTCATATGATCTAGGCAATGAACTCTGGACACATGACTGTTGTAAGAAG 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 464 TGGGAATGCATTCATACATCTGGGCAATGAACTCTGGGACATGACCATGACCATGGAAGAG 523

QY 297 TCATATCTTTGATGGTTCCTCTGAGAGTTCCAGCAATGGCAAGAAATCTCACTTATA 356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 TGATGCTCTTGAAGTTCCTCTGAGAGTTCCAGCAATGGCAAGAAATATCATATATA 583

QY 357 GGTCAAACTCTAATCAGAAATTCCTGATCTGTGACTGGACAGAGAGGCTGTGTTA 416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 AGTCAAACTCTAATTCAGAAATTCCTGATCTGTGACTGGACAGAGAGGCTGTGTTA 643

QY 417 CTGAAGTGAATATCCAGGTTCTTGTGTGCTTGTGGCTTTTCAGGCTCTGGGGGCC 476
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 CTGAAGTGAATATCAAGTTCCTTGTGTGCTTGTGGCTTTTCAGGCTCTGGGGGCC 703

QY 477 TGAAGCACTGAGTGAAGCTAAAGCAAGAAAGCTGGTGTCTCTGTGATCGACAGAACTTGG 536
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 TGAAGCACTGAGTGAAGCTGAAGCAAGAAAGCTGGTGTCTCTGTGATCGACAGAACTTGG 763

QY 537 TAGATTGCTCACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACAACTG 596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 764 TGAATTGCTCACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACAACTG 823

QY 597 CTTTCAGTATATTATGATAACAAAGGCAATGATTTCAGAAAGCTTCTCTATCCCTACAAAG 656
```

```
Db 824 CTTTCCAGTACATCATTGATAACAAGGCACTGCAGCTCTCTATCTCTATCTATCAAG 884
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 657 CCATGAATGGAAGTGCAGATATACATCAAAAAGGAGGCTGCACATGTTTAAATATA 716
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 CCATGATAGAAATATCAATATCA-ATAAATATGTTGTCACATATGTTAAATATA 944
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 717 CTGAACCTCCCTTGGCAGTGAAGATCACTTAAANAAGTGTGGCAATAAAGAGACTG 776
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 CTGAACCTCCCTTATGGSAGAGAGATGCTCTGAAGAGGCTGTGGGCAATAAAGAGCTCAG 1004
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 777 TGTCTGTGCTATAGATGAGAGCA-TATTCTTTCTCTCTATAGAAATGTTGTTACT 846
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 TGTCTGTGCTATAGATGAGAGCA-TATTCTTTCTCTCTATAGAAATGTTGTTACT 1064
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 837 ATGAACATCTCTACTCAGAAATGCTGAATCATGAGCAATTTAGCTTGGGCAATGCAACC 896
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1064 ATGAACATCTCTACTCAGAAATGCTGAATCATGAGCAATTTAGCTTGGGCAATGCAACC 1124
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 897 TTAATGGAAGACTACTGCTTGTGAAAAACAGTGGGCGCTCACTTTTGTGACACAG 956
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1124 TTAATGGAAGAAATACTGCTTGTGAAAAACAGTGGGCGCTCACTTTTGTGACACAG 1184
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 957 GATATATTGATGGAAGAAACAGTGGGCGCTCACTTTTGTGACACAGTCTT 1016
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1184 GATATATTGATGGAAGAAACAGTGGGCGCTCACTTTTGTGACACAGTCTT 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1017 ACCGAGAAATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAATTTGAACACACTTTC 1074
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1244 ACCGAGAAATCTAGAGATCTCTCTCTTTTATAACAATCAAGAAATGAGGACACTTTC 1404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTGCATGA 1122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1304 TCTTAATTTAATTTTCTGCTGTATCCAGAGAATAAATTTGTGTGCATGA 1353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-60-465-241-3608
; Sequence 3608, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3608
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-465-241-3608

Query Match 75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGACGACCTCGCAGGTTCTTTC---ATGAATTTGGCTAGTTGGGCTGC 57
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 GCACCTTACTCTRGAGCAACTGCTGCTCTATCACAATGAAGCGGCTGGTTGTGCGC 284

QY 58 TTCGTTGTGCTCTATGCAATGGGACAAAGTACATAAAGATCCCTCTGTGATCATCAT 117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 118 GGAATCTCTGGAAGAAACCTTACAGCAACAAT-CAAGGAAGAGAAATCAGGAAGTAGCAC 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 177 GCGTCTCATCTGGGAAAAAATCTAAATTTTGTGATGCTTCCACAACTCGAAGCACTCAA 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 237 TGGGAATGCATTTCATATGATCTAGGCAATGAACTCTGGACACATGACTGTTGTAAGAAG 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 297 TCATATCTTTGATGGTTCCTCTGAGAGTTCCAGCAATGGCAAGAAATCTCACTTATA 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 357 GGTCAAACTCTAATCAGAAATTCCTGATCTGTGACTGGACAGAGAGGCTGTGTTA 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 417 CTGAAGTGAATATCCAGGTTCTTGTGTGCTTGTGGCTTTTCAGGCTCTGGGGGCC 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 477 TGAAGCACTGAGTGAAGCTAAAGCAAGAAAGCTGGTGTCTCTGTGATCGACAGAACTTGG 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 537 TAGATTGCTCACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACAACTG 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 TCTTGGTGTGCTCTCTGCTGAGTGGGACAGATGTCATTAAGATCCCTACCCCTGGATCACCAC 117

QY 597 CTTTCAGTATATTATGATAACAAAGGCAATGATTTCAGAAAGCTTCTCTATCCCTACAAAG 246
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6918
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-466-412-6918

Query Match      75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTTATTAGTATGGGACCACTGCAGGTTCTTTC---ATGAAATGGCTAGTTGGGTGC 57
Db 224 GGACTCTTACTGTGGGCAACTGCTGTTCTATCACAATGAACGCGTGGTTGTGCGC 283

QY 58 TTCCGTTTGTCTCTATGCAGTGGCACAAGTACATAAAGATCCCACTCTCGATCATCACT 117
Db 284 TCTTGTGTGTCTCTCTGCTGAGTGGCAGATTCGATAAAGATCCCTACCTGGATCCACT 343

QY 118 GGAATCTCTGGAAGAAAACCTACAGCAAAAT-CAAGGAAGAGAAATGAGAAAGTATGAC 176
Db 344 GGCATCTCTGGAAGAAAACCTATGCAAAATATACAGGAAGAAATGAGAAAGTATGAC 403

QY 177 GCGTCTCTCATCTGGGAAAAAATCTAAATTTTGATGCTTCACAATCTGGAACACTCAA 236
Db 404 GAGGCTCATCTGGGAAAAAATCTAAATTTTGATGCTTCACAATCTGGAACACTCAA 463

QY 237 TGGGAATGCATTCATATGATCATTAGGCATCAACCACTCGGAGACATCACTCTCTCAACAAG 296
Db 464 TGGGAATGCATTCATGATCATTAGGCATCAACCACTCGGAGACATCACTCTCTCAACAAG 523

QY 297 TGATATCTTTGATGGGTTCCCTGAGAGTTCCAGCAATGCGAGAGAAATGTCTACTATA 356
Db 524 TGATATCTTTGATGGGTTCCCTGAGAGTTCCAGCAATGCGAGAGAAATGTCTACTATA 583

QY 357 GGTCAAACTCTAATCAGAAATTCCTGATTCCTGAGCTGGAGAGAGAGGCTCTGTATA 416
Db 584 AGTCAAACTCTAATGAGATTTGCTGATTCCTGAGCTGGAGAGAGAGGCTCTGTATA 643

QY 417 CTCAAGTGAATACCAAGGTTCTTGTGCTGCTTTGAGGCTTTTACGCGCTCTGAGGCGCC 476
Db 644 CTCAAGTGAATATCAAGGTTCTTGTGCTGCTTTGAGGCTTTTACGCGCTCTGAGGCGCC 703

QY 477 TGGAGACAACTGAGCTTAAACACAGAGAAAGTCTGCTCTGAGTGCACAGAACTTGG 536
Db 704 TGGAGACAACTGAGCTTAAACACAGAGAAAGTCTGCTCTGAGTGCACAGAACTTGG 763

QY 537 TAGATTGCTCAACTGAAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGCAACTG 596
Db 764 TAGATTGCTCAACTGAAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGCAACTG 823

QY 597 CTTTCCAGTATATTATGATAACACGGCAATTCATTCAGAGAGTTCTCTATCCCTACAAAG 656
Db 824 CTTTCCAGTATATTATGATAACACGGCAATTCATTCAGAGAGTTCTCTATCCCTACAAAG 883

QY 657 CCATGAATGAAAGTGCAGATATGACTCAAAAAGAGAGTGCATCATGTTCAAGATATA 716
Db 884 CCATGAATGAAAGTGCAGATATGACTCAAAAAGAGAGTGCATCATGTTCAAGATATA 943

QY 717 CTGAACCTTCCCTTTGGCAGTGAAGTCCCTTAAAGAAAGCTGTGGGCAATAAGGACCTG 776
Db 944 CTGAACCTTCCCTTTGGCAGTGAAGTCCCTTAAAGAAAGCTGTGGGCAATAAGGACCTG 1003

QY 777 TGTCTGTGGCTATAGATCGGAGCACTATTCTTCTTCTCTGATCAGAAAGTGTCTACT 836
Db 1004 TGTCTGTGGCTATAGATCGGAGCACTATTCTTCTTCTCTGATCAGAAAGTGTCTACT 1063

QY 837 ATGAACCTCTCTGACTCAGAAATGTGAATCATGAGTATTAAGTGTGCTATGTAACC 896
Db 1064 ATGAACCTCTCTGACTCAGAAATGTGAATCATGAGTATTAAGTGTGCTATGTAACC 1123

```

```

404 CACGTCATCTCGGAAAAAATCTAAAGTTTGTGATGCTTCAACACCTGGAGCAATCAA 463
QY 237 TGGGAATGCTATGATCTAGGCATGAACCATCTGGAGACATGACTGTGGAAGAG 296
Db 464 TGGGAATGCTATGATCTAGGCATGAACCATCTGGAGACATGACTGTGGAAGAG 523
QY 297 TGATATCTTTTGGGTTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATATA 356
Db 524 TGATATCTTTTGGGTTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATATA 583
QY 357 GGTCAAACTCTAATCAGAAATTCCTGATTCCTGAGCTGGAGAGAGAGGCTGTGTTA 416
Db 584 AGTCAAACTCTAATGAGATTTGCTGATTCCTGAGCTGGAGAGAGAGGCTGTGTTA 643
QY 417 CTGAAGTGAATATACAGGTTCTTGTGCTGCTTTGAGGCTTTTCAAGGCTGTGGGCGCC 476
Db 644 CTGAAGTGAATATACAGGTTCTTGTGCTGCTTTTCAAGGCTGTGGGCGCC 703
QY 477 TGGAGACAACTGAGCTTAAACACAGAGAAAGTCTGCTCTGAGTGCACAGAACTGG 536
Db 704 TGGAGACAACTGAGCTTAAACACAGAGAAAGTCTGCTCTGAGTGCACAGAACTGG 763
QY 537 TAGATTGCTCAACTGAAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGCAACTG 596
Db 764 TAGATTGCTCAACTGAAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGCAACTG 823
QY 597 TTTTCCAGTATATTATGATAACACGGCAATTCATGAGTCTCTGAGTGCACAGAACTG 656
Db 824 TTTTCCAGTATATTATGATAACACGGCAATTCATGAGTCTCTGAGTGCACAGAACTG 883
QY 657 CCATGAATGAAAGTGCAGATATGACTCAAAAAGAGAGTGCATCATGTTCAAGATATA 716
Db 884 CCATGAATGAAAGTGCAGATATGACTCAAAAAGAGAGTGCATCATGTTCAAGATATA 943
QY 717 CTGAACCTTCCCTTTGGCAGTGAAGTCCCTTAAAGAAAGCTGTGGGCAATAAGGACCTG 776
Db 944 CTGAACCTTCCCTTTGGCAGTGAAGTCCCTTAAAGAAAGCTGTGGGCAATAAGGACCTG 1003
QY 777 TGTCTGTGGCTATAGATCGGAGCACTATTCTTCTCTGATCAGAAAGTGTCTACT 836
Db 1004 TGTCTGTGGCTATAGATCGGAGCACTATTCTTCTCTGATCAGAAAGTGTCTACT 1063
QY 837 ATGAACCTCTCTGACTCAGAAATGTGAATCATGAGTATTAAGTGTGCTATGTAACC 896
Db 1064 ATGAACCTCTCTGACTCAGAAATGTGAATCATGAGTATTAAGTGTGCTATGTAACC 1123
QY 897 TTAATGGGAAGACTACTGCTGTGAAAGACAGCTGGGCGCTCAACTTTGGTGACCAAG 956
Db 1124 TTAATGGGAAGACTACTGCTGTGAAAGACAGCTGGGCGCTCAACTTTGGTGACCAAG 1183
QY 957 GATATATTCGGATGGGAAGACAGTGGAAATATATGCTGGGATTTGATTTACCTCTT 1016
Db 1184 GATATATTCGGATGGGAAGACAGTGGAAATATATGCTGGGATTTGATTTACCTCTT 1243
QY 1017 ACCAGAAATCTA---CAGCTCTCATTTTATTAACAGTCCCAAAAATTAAGCACTTTC 1073
Db 1244 ACCAGAAATCTAGAGATCTCTCTTTTATTAACAGTCCCAAAAATTAAGCACTTTC 1303
QY 1074 TCTTAATTTAATTTTACTGCTATA-ATAGTAAAAATTAAGTGTGATCA 1122
Db 1304 TCTTAATTTAATTTTCTGCTGTATCCAGAGAAATATGTTGTGATCA 1353

RESULT 9
US-60-466-412-6918
; Sequence 6918, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001466
; CURRENT APPLICATION NUMBER: US/60466.412

```

```

QY 897 TTAATGGGAAAGACTACTGGCTTGTAAGAAACAGCTGGGGCTCAACITTTGGTGACCAAG 956
DB 1124 TTAATGGGAAAGAACTACTGGCTTGTAAGAAACAGCTGGGGCTCAACITTTGGTGAAAG 1183
QY 957 GATATATTCGGATGGCAAGAAACAGCTGGGAAATCACTTGGGATGCTAGTTATCCCTCTT 1016
DB 1184 GATATATTCGGATGGCAAGAAATCAAGGAAATCACTTGGGATGCTAGTTATCCCTCTT 1243
QY 1017 ACCCAAAATCTA---CAGCTCTTCATTTTATAACAAAGTCCAAAAATTTGAACACTTTC 1073
DB 1244 ACCCAAAATCTAGAGGATCTCTCTTTTATAACAAATCAAGAAATTTGAACACTTTC 1303
QY 1074 TCTTAATTTAAATTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 1122
DB 1304 TCTTAATTTAAATTTCTCTCTCTATCCAGAAATAAATTTGTGTCATGA 1353

```

RESULT 10

```

US-09-873-367C-173
; Sequence 173, Application US/09873367C
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meena
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290 64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 173
; LENGTH: 1763
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-173

```

```

Query Match 75.1% Score 845.4 DB 5: Length 1763;
Best Local Similarity 87.2% Pred.No. 9.5e-218.
Matches 986; Conservative 0; Mismatches 136; Indels 9; Gaps 5;

```

```

QY 1 GCATTTATAGTATGGGAGCACTGCGAGCTCTTTC---ATCAATGGCTAGTTGGCGTCC 57
DB 99 GGACTCTTACTGTAGGCACTGCTGGTCTCTATCACAATGAACGCTGGTTTGTGTC 158
QY 58 TTCCGTTGCTCTATGCACTGGGACAGTACATTAACATCCCACTCTGGATCATCAT 117
DB 159 TCTTGTGCTGCTCTGCACTGGGACAGTCTGCATTAAGATCCCTACCCCTGGATCACCAC 218
QY 118 GGAATCTCTGGAAGAAAGCTACAGCAACAAT-CAAGAGAGAAATGAGGAGTAGCAC 176
DB 219 GGCATCTCTGGAAGAAAGCTATGCAACAATACAGGAAAGAAATGAGGAGTAGCAC 278
QY 177 GCGCTCTATCTGGGAAAAAAATCTAAATTTGTGATGCTTACAACTCTGGAACACTCAA 236
DB 279 GAGCTCTATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCAAACTTGGAGCATTCAA 338
QY 237 TGGGAATGATTCATATGATCTAGGCATCAACCTCTGGGACACATGCTGCTGAAGAAG 296
DB 339 TGGGAATGATTCATACGATCTGGGACATCAACCTCTGGGACACATGACCATGGAAGAAG 398
QY 297 TATATCTTTTATAGGTTTCTAGAGTTCTCCAGAGTCCAGAGAAATGTCTACTTATA 356

```

```

DB 399 TGATGTCTTTCATGAGTTCTTGAGAGTTCTTGAAGCTGGCAGAGAAATATACATATA 458
QY 357 GGTCAAACTCTAATCAGAAATTTGCTGATCTTGCGACTTGGAGAGAGAGAGAGAGAGAG 416
DB 459 AGTCAAAAGCTTAAATGAGATATTGCTATTCTTGGA-TGGAGAGAGAGAGAGAGAGAG 518
QY 417 CTGAAGTCAAAATACAGAGTTCTTGTGATCTTGTGGATTTTGTGGATTTTGTGGAGGCC 476
DB 519 CTGAAGTCAAAATACAGAGTTCTTGTGATCTTGTGGATTTTGTGGATTTTGTGGAGGCC 578
QY 477 TGAAGGACAACTCAAGCTAAAAACAGGAAAGCTGGTCTCTGAGTGCACAAAGAACTTGG 546
DB 579 TGGAGGACAACTCAAGCTAAAAACAGGAAAGCTGGTCTCTGAGTGCACAAAGAACTTGG 648
QY 537 TAGATTTGCTCAACTCAAGCTAAAAACAGGAAAGCTGGTCTCTGAGTGCACAAAGAACTTGG 596
DB 639 TGGATTTGCTCAACTCAAGCTAAAAACAGGAAAGCTGGTCTCTGAGTGCACAAAGAACTTGG 698
QY 597 CTTTCCAGTATATTATTGATTAACAGGCTTATTGATTGAGAGCTTCTCTATCTATCAAAAG 656
DB 699 CTTTCCAGTATATTATTGATTAACAGGCTTATTGATTGAGAGCTTCTCTATCTATCAAAAG 758
QY 657 CCATGAATGSAAGTGCAGATATGACTCAAAAAAGAGCTGGCAGCTTCTTCAAAAGTATA 716
DB 759 CCATGAATGSAAGTGCAGATATGACTCAAAAAAGAGCTGGCAGCTTCTTCAAAAGTATA 818
QY 717 CTGAAGTCAAAATACAGAGTTCTTGTGATCTTGTGGATTTTGTGGATTTTGTGGAGGCC 776
DB 819 CTGAAGTCAAAATACAGAGTTCTTGTGATCTTGTGGATTTTGTGGATTTTGTGGAGGCC 878
QY 777 TGTCTGTGGCTATAGTATGCGAGCCACTTATCTTCTCTCTGATCAAGAGTGTCTACT 836
DB 879 TGTCTGTGGCTATAGTATGCGAGCCACTTATCTTCTCTCTGATCAAGAGTGTCTACT 938
QY 837 ATGAACCATCTTGTATCTAGAAATGTGAATCATGAGTATTAGTATAGTATAGTATAGT 896
DB 939 ATGAACCATCTTGTATCTAGAAATGTGAATCATGAGTATTAGTATAGTATAGTATAGT 998
QY 897 TTAATGGGAAAGACTACTGGCTTGTAAGAAACAGCTGGGGCTCAACITTTGGTGACCAAG 956
DB 999 TTAATGGGAAAGAACTACTGGCTTGTAAGAAACAGCTGGGGCTCAACITTTGGTGACCAAG 1058
QY 957 GATATATTCGGATGGCAAGAAATCAAGGAAATCACTTGGGATGCTAGTTATCCCTCTT 1016
DB 1059 GATATATTCGGATGGCAAGAAATCAAGGAAATCACTTGGGATGCTAGTTATCCCTCTT 1118
QY 1017 ACCCAAAATCTA---CAGCTCTTCATTTTATAACAAAGTCCAAAAATTTGAACACTTTC 1072
DB 1119 ACCCAAAATCTAGAGGATCTCTCTTTTATAACAAATCAAGAAATTTGAACACTTTC 1178
QY 1073 CTCTTAATTTAAATTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 1122
DB 1179 CTCTTAATTTAAATTTCTCTCTCTATCCAGAAATAAATTTGTGTCATGA 1229

```

RESULT 11

```

US-10-219-051H-2756
; Sequence 2756, Application US/10219051H
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: ICA 35693 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/219,051H
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715

```



```

935 GGCCTCAACTTTGGTGAAGGATATATCGGATGGCAAGAACAGTGGAAATCACTGT 994
1007 GGCCTCAACTTTGGTGAAGGATATATCGGATGGCAAGAACAGTGGAAATCACTGT 996
995 GGCCTCAACTTTGGTGAAGGATATATCGGATGGCAAGAACAGTGGAAATCACTGT 1051
996 GGCCTCAACTTTGGTGAAGGATATATCGGATGGCAAGAACAGTGGAAATCACTGT 1026
1052 TCCAAAAATTTGAACACATTTCTCTTAATTTTACCTGCTATATA-ATAGTAAATATA 1110
1027 TCAAGAAATATGAACACATTTCTCTTAATTTTACCTGCTATATA-ATAGTAAATATA 1086
1111 AATGTGTCATGA 1122
1087 ATTGTGTCATGA 1098

RESULT 14
US-09-724-676-33122
; Sequence 33122, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33122
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-33122

Query Match 73.3%; Score 824.8; DB 6; Length 1366;
Best Local Similarity 86.8%; Pred. No. 3.3e-212;
Matches 944; Conservative 0; Mismatches 137; Indels 7; Gaps 3;

1 GCATTATTAGTATGGAGCACTGCAGGTTCTTTC---ATGAAATGGCTAGTTGGGCTGC 57
99 GCATTATTAGTATGGAGCACTGCAGGTTCTTTC---ATGAAATGGCTAGTTGGGCTGC 158
58 TTCCGTTGTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 117
159 TCTTGTGCTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 218
118 GGAATCTTGGAGAAACCTACAGCAACAT-CAAGGAGAGAAATGAGGAAGTACAC 176
219 GGCATCTCTGAGAAACCTATGAGCAACATCAAGGAAGTACACAGGAGTAC 278
177 GGCCTCTCATCTGGGAAAAATCTAAAAATTTGTGATGCTTCAACATCTGGAACACTCA 236
279 GAGCTCTCATCTGGGAAAAATCTAAAAATTTGTGATGCTTCAACATCTGGAACACTCA 338
237 TGGAAATGCTATATGATCTAGGATGAGCACTCTGGGAGACATGACTGGTGAAGAG 296
339 TGGAAATGCTATATGATCTAGGATGAGCACTCTGGGAGACATGACTGGTGAAGAG 398
297 TGATATCTTTGATGGGTTCCCTGAGAGTTCAGGATCCAGCAATGGGAGCAATGTCATTATA 356
399 TGATGCTTTGATGGGTTCCCTGAGAGTTCAGGATCCAGCAATGGGAGCAATGTCATTATA 458
357 GGTCAAACTCTAATCAAAATTCCTGATCTGAGTGGAGTGGAGAGGAGGCTGTGTTA 416
459 AGTCAAACTCTAATCAAAATTCCTGATCTGAGTGGAGTGGAGAGGAGGCTGTGTTA 518
417 CTGAAGTGAATACAGGGTCTTGTGCTGCTTGTGGGCTTTCAGGCTGTGGGGGCC 476
519 CTGAAGTGAATACAGGGTCTTGTGCTGCTTGTGGGCTTTCAGGCTGTGGGGGCC 578
477 TGGAGGACAACTGAAGCTTAAACAGAGAAAGCTGCTGCTGAGTGGAGCAAGCTTGG 536
579 TGGAGGACAACTGAAGCTTAAACAGAGAAAGCTGCTGCTGAGTGGAGCAAGCTTGG 638

```

```

537 TAGATTGCTCAACTGAAAAATATGGAAATAAAGGCTGATGGGGTTTTCATGCAACTG 596
639 TGGATTGCTCAACTGAAAAATATGGAAATAAAGGCTGATGGGGTTTTCATGCAACTG 698
597 CTTTCAGTATATTATGATACAAAGCATTGATTGATGAGAGCTTCTCTATCCCTACAAG 656
699 CTTTCAGTATATTATGATACAAAGCATTGATTGATGAGAGCTTCTCTATCCCTACAAG 758
657 CCATGAATGGAAGTGCAGATATGACTCAAAAAAGCAGCTGGCCACATGTTCAAAAGTATA 716
759 CCATGAATGGAAGTGCAGATATGACTCAAAAAAGCAGCTGGCCACATGTTCAAAAGTATA 818
717 CTGAACCTCCCTTTGGCAGTGAAGTGCCTTAAAAAGAGCTGGCCCAATAAAGGACCTG 776
819 CTGAACCTCCCTTTGGCAGTGAAGTGCCTTAAAAAGAGCTGGCCCAATAAAGGACCTG 878
777 TGTCTGTGCTATAGATGGAGCCATATCTTTCTCTGTACAGAGTGTGTCTACT 836
879 TGTCTGTGCTATAGATGGAGCCATATCTTTCTCTGTACAGAGTGTGTCTACT 938
837 ATGAACCATCTCTACTCAGAAATGGAATCATGAGTATTAGTGGTTGGCTATGTTAACC 896
939 ATGAACCATCTCTACTCAGAAATGGAATCATGAGTATTAGTGGTTGGCTATGTTAACC 998
897 TTAATGGGAAAGACTACTGGCTTGTGAAAAAGCAGCTGGGCGCTCAACTTTGGTGACCAAG 956
999 TTAATGGGAAAGACTACTGGCTTGTGAAAAAGCAGCTGGGCGCTCAACTTTGGTGACCAAG 1058
957 GATATATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATGCTAGTATTATCCCTCTT 1016
1059 GATATATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATGCTAGTATTATCCCTCTT 1118
1017 ACCCAGAAATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAATTCAAACACTTTC 1073
1119 ACCCAGAAATCTAGAGGATCTCTCTTTTATAACAAGTCCAAAAATTCAAACACTTTC 1178
1074 TCTTAATT 1081
1179 GTTTCATT 1186

RESULT 15
US-09-724-676A-33122
; Sequence 33122, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33122
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-33122

Query Match 73.3%; Score 824.8; DB 6; Length 1366;
Best Local Similarity 86.8%; Pred. No. 3.3e-212;
Matches 944; Conservative 0; Mismatches 137; Indels 7; Gaps 3;

1 GCATTATTAGTATGGAGCACTGCAGGTTCTTTC---ATGAAATGGCTAGTTGGGCTGC 57
99 GCATTATTAGTATGGAGCACTGCAGGTTCTTTC---ATGAAATGGCTAGTTGGGCTGC 158
58 TTCCGTTGTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 117
159 TCTTGTGCTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 218
58 TTCCGTTGTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 117
159 TCTTGTGCTGCTCTTATGCTGAGTGGCAAGTACATTAAGATCCCACTCTGATCATCACT 218
118 GGAATCTTGGAGAAACCTACAGCAACAT-CAAGGAGAGAAATGAGGAAGTACAC 176
219 GGCATCTCTGAGAAACCTATGAGCAACATCAAGGAAGTACACAGGAGTAC 278
177 GGCCTCTCATCTGGGAAAAATCTAAAAATTTGTGATGCTTCAACATCTGGAACACTCA 236
279 GAGCTCTCATCTGGGAAAAATCTAAAAATTTGTGATGCTTCAACATCTGGAACACTCA 338
237 TGGAAATGCTATATGATCTAGGATGAGCACTCTGGGAGACATGACTGGTGAAGAG 296
339 TGGAAATGCTATATGATCTAGGATGAGCACTCTGGGAGACATGACTGGTGAAGAG 398
297 TGATATCTTTGATGGGTTCCCTGAGAGTTCAGGATCCAGCAATGGGAGCAATGTCATTATA 356
399 TGATGCTTTGATGGGTTCCCTGAGAGTTCAGGATCCAGCAATGGGAGCAATGTCATTATA 458
357 GGTCAAACTCTAATCAAAATTCCTGATCTGAGTGGAGTGGAGAGGAGGCTGTGTTA 416
459 AGTCAAACTCTAATCAAAATTCCTGATCTGAGTGGAGTGGAGAGGAGGCTGTGTTA 518
417 CTGAAGTGAATACAGGGTCTTGTGCTGCTTGTGGGCTTTCAGGCTGTGGGGGCC 476
519 CTGAAGTGAATACAGGGTCTTGTGCTGCTTGTGGGCTTTCAGGCTGTGGGGGCC 578
477 TGGAGGACAACTGAAGCTTAAACAGAGAAAGCTGCTGCTGAGTGGAGCAAGCTTGG 536
579 TGGAGGACAACTGAAGCTTAAACAGAGAAAGCTGCTGCTGAGTGGAGCAAGCTTGG 638

```

Db 219 GGCACTCTGGAGAAAGCACTATGGTAAACAATAAAGGAAAAAATGAAGAGCAGTAC 278
QY 177 GGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACAATCTGGAACTCAA 236
Db 279 GAGCTCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCACAACCTGGAGCATCAA 338
QY 237 TGGGAATGCAATCATATGATCTAGGCATGAACCATCTGGGAGACATGACTGGTGAAGAG 296
Db 339 TGGGAATGCATCATACGATCTGGGCATGAACCATCTGGGAGACATGACCATGAAGAG 398
QY 297 TGATATCTTTGATGGTTCCCTGACAGTTCCAGGCCAATGGCAGAGAAATGTCACCTATA 356
Db 399 TGATGCTTTGATGATTTCCCTGAGAGTTCCAGCAAGTCCAGCAAGTTCACATATA 458
QY 357 GGTCAAACTCTAATCAGAAATTCCTGATTTCTGCTGACCTGGAGAGAGAGGGCTGTGTTA 416
Db 459 AGTCAAACTCTAATCGGATATTCCTGATTTCTGCTGACCTGGAGAGAGAGGGTGTGTTA 518
QY 417 CTGAAGTGAATATACCAAGTTCTTGTGTTGCTTTGTTGGCTTTTCAGCGCTTGGGGGCC 476
Db 519 CTGAAGTGAATATCAAGCTTCTTGTGCTGCTTGTGCTGCTTTCAGTGTCTGGGGGCC 578
QY 477 TCGAAGCACAACCTGAAGCTTAAACACAGGAAGCTGCTGCTCTGACTGCACAGAACTTGG 536
Db 579 TGGAGCAGAGTGAAGCTGAAGCTGAAGCAGGAAGCTGCTGCTGACTGCACAGAACTTGG 638
QY 537 TAGATTTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTTCATCAACAAG 596
Db 639 TGGATTTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCTTCATCAACAAG 656
QY 597 CTTTCAGTATATATTGATACAGGGCATTTGATTCAGAGCTTCCCTATCCCTACAAG 656
Db 699 CTTTCAGTATACATCATATGATACAGGGCATTCGACTCAGACGCTTCCCTATCCCTACAAG 758
QY 657 CCATGAATGGAAGTGCATATGACTCAAAAAAGGAGCTGCACATCTTCCAAAGTATA 716
Db 759 CCATGAATGGAAGTGCATATGACTCAAAAAAGGAGCTGCACATCTTCCAAAGTATA 818
QY 717 CTGAACCTCCCTTTGGCAGTGAAGATGGCTTAAAGAGAGCTGGCCCAATAAAGAGCCTG 776
Db 819 CTGAACCTCCCTTTATGCCAGAGAGATGCTCTCAAAAGAGCTGGCCCAATAAAGAGCCTG 878
QY 777 TGTCTGTGCTATAGATGGAGCCACTATTCTTCTCTGTACAGAGTGGTGTCTACT 836
Db 879 TGTCTGTGCTATAGATGGAGCCACTATTCTTCTCTGTACAGAGTGGTGTCTACT 938
QY 837 ATGAACCATCTCTACTCAGATGTGAATCATGAGTATTTAGTGGTGGCTATGCTAAACC 896
Db 939 ATGAACCATCTCTACTCAGATGTGAATCATGAGTATTTAGTGGTGGCTATGCTAAACC 998
QY 897 TTAATGGGAAGCACTACTGGCTTTGGAAGAGAGCTGGGGCTCAACTTTGGTGACCAAG 956
Db 999 TTAATGGGAAGCACTACTGGCTTTGGAAGAGAGCTGGGGCTCAACTTTGGTGACCAAG 1058
QY 957 GATATATTCGGATGGCAAGAAAGTGGAAATCACTGTGGGATTTGCTAGTTATCCCTCTT 1016
Db 1059 GATATATTCGGATGGCAAGAAATGAAGAAATCATTGTGGGATTTGCTAGCTTCCCTCTT 1118
QY 1017 ACCCAGAAATCTA---GACCTCTCATTTTATAAGAGTCCAAAAAATTTGAACACTTTC 1073
Db 1119 ACCCAGAAATCTAGAGGATCTCTCTTTTATAACAAATCAAGAAATATAGAAAAATATA 1178
QY 1074 TCTTAATT 1081
Db 1179 GTTTGATT 1186

Search completed: July 1, 2003, 12:00:03
Job time : 989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:	June 27, 2003, 11:05:43 ; Search time 72 Seconds (without alignments) 4167.783 Million cell
---------	---

Title: US-10-010-580-1
 Perfect score: 1997
 Sequence: 1gcattattagtatggagca.....ataaataatgtctatgacct 1126

Scoring table:	BLOSUM62	
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

Searched: 908470 seqs, 133250620 residues

total number of hits	satisfying	chosen	parameters:
1816940			

Minimum DH seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 s

Command line parameters:

```

MODELS=linnet_r2p_model DEV=xlp
MODELS+=(USPTL_scool US10010580 FUNAL_27062003 104535 11686/app_query.fasta_1.1387
-DNA_COSSES 101002 OFMT=fastan SUFFIX=n2p tag MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=1 UNIS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -LOCALCUT=200 -THR_SCORE=exp -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL OUTFM=p2o -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USE=ts10010580 ALIGN=1 -L114 FUNAL_27062003 104535 11686 -NCPU=3
-NO_MAP -LARGESTORY -NEG_SCORE=0 -WAIT -DSPBLCK=100 -LONGLOG
DEV TIMOUT=120 -WARN TIMOUT=30 -THRAUD=1 -XCAPEXT=10 -XGAPEXT=0.5 -FGAPE=6
-DELEV=7 -VCAPBP=10 -YGAPEXT=0.5 -DELDP=7

```

Database : A_Geneseq_101002.*

- | | |
|----|---|
| 1 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1980.DAT.* |
| 2 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1981.DAT.* |
| 3 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1982.DAT.* |
| 4 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1983.DAT.* |
| 5 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1984.DAT.* |
| 6 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1985.DAT.* |
| 7 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1986.DAT.* |
| 8 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1987.DAT.* |
| 9 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1988.DAT.* |
| 10 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1989.DAT.* |
| 11 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1990.DAT.* |
| 12 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1991.DAT.* |
| 13 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1992.DAT.* |
| 14 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1993.DAT.* |
| 15 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1994.DAT.* |
| 16 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1995.DAT.* |
| 17 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1996.DAT.* |
| 18 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1997.DAT.* |
| 19 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1998.DAT.* |
| 20 | /SID52/qcqdta/geneseq/geneseq-emb1/AA1999.DAT.* |
| 21 | /SID52/qcqdta/geneseq/geneseq-emb1/AA2000.DAT.* |
| 22 | /SID52/qcqdta/geneseq/geneseq-emb1/AA2001.DAT.* |
| 23 | /SID52/qcqdta/geneseq/geneseq-emb1/AA2002.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1569	78.6	331	21	AA559634	Human Cathepsin S	Human Cathepsin S
2	940.5	47.1	329	21	AA830506	A monkey cathepsin	A monkey cathepsin
3	939.5	47.0	329	16	AA882720	Human osteocalcin-d	Human osteocalcin-d
4	939.5	47.0	329	17	AA895599	Prepro-cathepsin-D	Prepro-cathepsin-D
5	939.5	47.0	329	19	AAW59416	Human cathepsin K	Human cathepsin K
6	939.5	47.0	329	19	AAW41645	Human cathepsin K	Human cathepsin K
7	939.5	47.0	329	21	AAV81000	Human procathepsin	Human procathepsin
8	939.5	47.0	361	21	AAH57115	Human prostate can	Human prostate can
9	937.5	46.9	329	19	AAW77071	Rat Cathepsin K po	Rat Cathepsin K po
10	932.5	46.7	329	23	AAW41984	Human cathepsin X	Human cathepsin X
11	932.5	46.7	329	23	ABR81810	Human cathepsin X	Human cathepsin X
12	930.5	46.6	329	19	AAW48947	Mutant human cathe	Mutant human cathe
13	930.5	46.6	329	19	AAW48939	Mutant human cathe	Mutant human cathe
14	929.5	46.5	329	19	AAW48946	Mutant human cathe	Mutant human cathe
15	924.5	46.5	329	19	AAW48938	Mus musculus cathe	Mus musculus cathe
16	925.5	46.3	329	19	AAW42102	Mouse ischaemic co	Mouse ischaemic co
17	922.5	46.2	329	23	ABH57267	Human recombinant	Human recombinant
18	915	45.8	329	19	AAW44778	Human recombinant	Human recombinant
19	901	45.1	322	22	AAH66260	polypeptide identi	polypeptide identi
20	843	42.2	334	22	AAV02358	Human cathepsin L2	Human cathepsin L2
21	843	42.2	334	21	AAV81487	Human cathepsin L2	Human cathepsin L2
22	843	42.2	334	22	AAU12177	Human PR0305 polyp	Human PR0305 polyp
23	832	41.7	334	22	AAW94300	Human cathepsin (L	Human cathepsin (L
24	812.5	40.7	336	22	ABG21426	Novel human diapo	Novel human diapo
25	808	40.5	333	22	AAW53531	Human polypeptide	Human polypeptide
26	806	40.4	333	19	AAW47031	Human procathepsin	Human procathepsin
27	806	40.4	333	23	ABW77396	Human cathepsin L	Human cathepsin L
28	805.5	40.3	333	23	ABG66692	Human novel polype	Human novel polype
29	798.5	40.0	333	23	ABG66714	Human novel polype	Human novel polype
30	793	38.7	333	13	AAK28829	Human procathepsin	Human procathepsin
31	792	39.7	333	9	AAK82094	Phu-16 sequence on	Phu-16 sequence on
32	767	38.4	341	22	ABG62597	Drosophila melanog	Drosophila melanog
33	738.5	37.0	334	23	AAU98883	Human proctase PRT	Human proctase PRT
34	732	36.2	333	23	ABG61771	Novel cathepsin-L	Novel cathepsin-L
35	709.5	35.5	326	15	AAW57071	Fasciola hepatica	Fasciola hepatica
36	705.5	35.3	419	21	AAK35414	Zea mays protein f	Zea mays protein f
37	705	35.3	424	21	AAK35413	Zea mays protein f	Zea mays protein f
38	701	35.1	330	23	AAH09882	Amino acid sequenc	Amino acid sequenc
39	689.5	34.5	461	20	AAW89558	Triticum sp. cyste	Triticum sp. cyste
40	686.5	34.4	326	15	AAK57080	Fasciola hepatica	Fasciola hepatica
41	681	34.1	330	21	AAK13420	Marine sponge sili	Marine sponge sili
42	665	33.3	211	19	AAW44779	Human cathepsin K	Human cathepsin K
43	651	32.6	472	20	AAW89559	Triticum sp. cyste	Triticum sp. cyste
44	649.5	32.5	500	24	AAH65766	Cysteine protease	Cysteine protease
45	649	32.5	463	18	AAW19541	Soybean thiol prot	Soybean thiol prot

ATTACHMENTS

RESULT 1	
AAV59634	
ID	AAV59634 standard; Protein; 331 AA.
XX	
XX	AAV59634;
XX	
XX	27-MAR-2000 (first entry)
DT	
XX	
XX	Human Cathepsin S amino acid sequence.
DE	
XX	Cathepsin S; human; endometriosis; treatment; diagnose.
XX	
XX	
XX	Homo sapiens.
XX	OS
XX	WC0963115-A2.
PN	
XX	
XX	09-DEC-1999.
XX	
XX	03-JUN-1999; 99WO-US12335.
PF	

US-10-010-580-1 (1-1126) x AAW39216 (1-329)

Alignment Scores:

Alignment Scores:
pred. No.: 3,350-95 length: 329

Score: 939.50 Matches: 183
 Percent Similarity: 70.438 Conservative: 48
 Best Local Similarity: 55.798 Mismatches: 88
 Query Match: 47.05% Indels: 10
 DB: 21 Gaps: 4

US-10-010-580-1 (1-1126) x AAM57115 (1-361)

```

QY 54 CTGCTCCGCTGTGCTCTATGAGTGGACAAAGTACATMAAGATCCACTGTGATCAT 113
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 40 LeuLeuProValValSerPheLea-----LeuTyrProGluGluLeuLeuSptThr 56
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 114 CACTGCAATCTGCGAAGAAACCTACAGCAAAACATCAAGACAGACAT--GAGCAAGTA 172
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 57 HISTRGluLeuTyrPheLysThrHisAlaGlySerGlnTyrAsnAsnLysValAspLeuIle 76
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 173 GTCAGCGGCTGTCATCTGAGAAAAAAATTTAAATTTGTGATGCTTCACAACTGGACAC 232
    .....|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 77 SerArgArgLeuLeuIleProGluLysAsnLeuLysTyrIleSerIleHisAsnLeuGluAla 96
    .....|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 233 TTTAATGGAAATTCATTCATATATATATATGAGTGAATCATCTGGAGCATGCTGTTGAA 292
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 97 SerLeuGlyValHisThrTyrGluLeuAlaMetAsnHisLeuGlyAspMetThrSerGlu 116
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 293 GAAGTGAATCTTGTGATGGCTCCCTGACAGATCCAGCCCAATGGCAGACA-----AAT 346
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 117 GluValValGlnLysMetThrGlyLeuLysValProLeuSerHisSerArgSerAsnAsp 136
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 347 GTGACTTATATAGTCAAACTATATCAAAATTCGCTGATTCGTGTGACTGGAGAGAGAG 406
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 137 ThrLeuTyrIleProGluIleProGluIleArgGluAlaProAspSerValAspTyrArgLysLys 156
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 407 GGCTGTGTTGATGAGTGAATTCAGAGGTTCTTGTGGTCTGTTGGCTTTTCACAGCT 466
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 157 GlyTyrValThrProValLysAsnGlnGlyGlnGlySerGlyStrAlaPheSerSer 176
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 467 GTGGGGCCCTGGAACACACTGGAAGCTAAAAACAGCAAGCTGCTGCTCTGACATGCA 526
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 177 ValGlyAlaLeuGlnLysGlnLeuLysLysThrGlyLysLeuLeuAsnLeuSerPro 196
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 527 CCAAACTTGTGATGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 586
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 197 GluAsnLeuValAspCysValSerGlu-----AsnAspGlySerGlyGlyGlyTyr 213
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 587 ATGCAAACTGCTTCCAGTATATATATATATATATATATATATATATATATATATATAT 646
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 214 MetThrAsnAlaPheSerGlnTyrValGlnLysAsnArgGlyIleAspSerGluAspAlaTyr 233
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 647 CCTTACAAAGCCATGGAATGGAAGCTGCAGATATACATCAAAAGAGAGCTGCCACATGT 706
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 234 ProTyrValGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 253
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 707 TCAAGATATATATATATATATATATATATATATATATATATATATATATATATATATAT 766
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 254 ArgGlyTyrArgGlnIleProGluGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 273
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 767 AAAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 826
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 274 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheLeuIlePheTyrSerLys 293
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 827 GGTGTCTACATGAGCAAGCTCTG---ACACAGATCTGATCATGATGATATAGTGTGTGT 883
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 294 GluValAlaTyrTyrArgLysSerCysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 313
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 884 GATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 943
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 314 GluTyrGlyIleGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 333
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 944 TTTGTGACCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1003
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 334 TrpLysAsnLysGlyTyrIleLeuMetAlaIleArgAsnLysAsnAlaCysGlyIleAla 353
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 1004 AGTTATCCCTCTTACCCCAAAATC 1027
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||

```

Db 354 AsnLeuAlaSerPheProLysMet 361

RESULT 9

AAM77071

AAM77071 standard; Protein: 329 AA.

09-NOV-1998 (first entry)

Rat Cathepsin K polypeptide.

Rat; cathepsin K; cysteine protease; osteoclast; bone resorption;

osteoporosis; periodontal disease; arthritis; cancer; tumour metastasis;

Paget's disease.

Rattus sp.

EP861898-A1.

02-SEP-1998.

23-DEC-1997; 97EP-0310523.

26-FEB-1997; 97US-0806959.

(SMIK) SMITHKLINE BEECHAM CORP.

Brun K, Feld JA;

WPI; 1998-449110/39.

N-PSDB; AAV48225.

DNA encoding rat cathepsin K polypeptide - and corresponding e.g.

polypeptide, antibody, agonist and antagonist etc.

Claim 15; Fig 1; 19pp; English.

The Rat cathepsin K is a novel cysteine protease found to be abundantly

and selectively expressed in osteoclasts, which suggests a role in bone

resorption. The cathepsin K polypeptides and polynucleotides can be used

to produce agonists and antagonists to treat e.g. osteoporosis,

periodontal disease, arthritis, cancer, tumour metastasis and Paget's

disease.

Sequence 329 AA:

US-10-010-580-1 (1-1126) x AAM77071 (1-329)

```

QY 54 CTGCTCCGCTGTGCTCTATGAGTGGACAAAGTACATMAAGATCCACTGTGATCAT 113
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 8 LeuLeuProValValSerPheLea-----ProGluGluLeuLeuSptThr 24
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 114 CACTGCAATCTGCGAAGAAACCTAAGAAATTTGTGATGCTTCACAAATCTGCAACAC 232
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 25 SerArgArgLeuLeuIleProGluLysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 44
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 173 GTCAGCGGCTGTCATCTGAGAAAAAAATTTAAATTTGTGATGCTTCACAACTGGACAC 232
    .....|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 45 SerArgArgLeuLeuIleProGluLysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 64
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
QY 233 TCAATGGAATGTCATATATATATATATATATATATATATATATATATATATATATATATATAT 292
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||
Db 65 SerLeuGlyAlaHisThrTyrGlnLeuAlaMetAsnHisLeuGlyAspMetThrSerGlu 84
    |||||...|||...|||...|||...|||...|||...|||...|||...|||...|||

```

Alignment Scores: 5.59c-95 Length: 329

Score: 937.50 Matches: 179

Percent Similarity: 70.73% Conservative: 53

Best Local Similarity: 54.57% Mismatches: 87

Query Match: 46.95% Indels: 10

DB: 19 Gaps: 4

QY		527	CACAACTTCGTGATAGTTCACACTGCAGAAAATATGGCAATAAAGGCTGCAATGGCGCTTTC	586
Db		165	GlnAsnLeuValAlaIspcysValSerGlu-----AsnAspGlyCysGlyGlyGlyIyr	181
QY		587	ATACAAATGATTTCCTTAATATTATTATATGAACAAGCATTTGATTCAGAAGCTTCAT	646
Db		182	MetThrAsnAlaIheGlnIyryValGlnGlyAsnArgrGlyIleAspSerCIAspAlaIyrr	201
QY		647	CCCTACAAAGCCATGAAATGGAAAAAGTGCAATATGACTCAA AAAAGCGAGTGGCCACATGT	706
Db		202	PrcItyrValGlyGlnIoclucIusercysMetTyraSprProthrGlyLysAlaIalyscys	221
QY		707	TCAAGATATACCTAACCTCCCTTGGCACGCAAGATGCCCTAAAYAGACCTGTGGCCAAT	766
Db		222	ArgGlyTyraIrgclInilePrcIocIugIyasnGlnIySalaleuIyAdgAlaValAlaIarg	241
QY		767	AAAGGACCTGTGTCTGTGGCTATAGATGCGAGACACTATCTTCTCTCTGTACGAAGT	826
Db		242	ValclIyProvalIserValAlaIleAspAlaIserLeuthIserPhcIinIhetTySerlys	261
QY		827	GGTGTCTACTATGAAACATCTGT--ACTCGAAGATGTGAATCAAGATATTAGTGGTT	883
Db		262	GlyValItyrTrtyrasprIunserCysAsnsenrAspAsnleuAsnhIalaValIeuaIval	281
QY		884	GGCTATGTACCTTAATGGGAAAGACTACTGGCTTGTGAAAACAGCTGGGGCTCAAC	943
Db		282	GlyIyrgIylIleGlnIySgIyasnIyShIstrIpIleIleIyasnSerIrpIyGluasn	301
QY		944	TTTGCTAGCCAGAGATATATTGGATGCGTAAAGAAATATGAAATPACTGTGGATGTCT	1003
Db		302	IrpIyasnIySgIyIyrrIleleuMetaIalaIrgasnlIyAsnasnAlacysglIyleala	321
QY		1004	AGTTATCCCTCTTACCAGAAATC 1027	
Db		322	AsnleuAlaserPhePrPolysmet 329	
RESULT 11				
AAB81810				
ID	AAB81810	standard:	Protein: 329 AA.	
XX	AC	AAB81810;		
XX	DT	16-SEP-2002 (first entry)		
XX	DE	Human cathepsin X.		
XX	KW	Human: osteoclast; gene therapy; aberrant bone resorption;		
XX	OS	Homo sapiens.		
PN	US6403304-B1			
XX	PD	11-JUN-2002.		
XX	PF	19-JUL-1996;	96US-0684932.	
XX	PR	26-MTE-1995;	95US-001292P	
XX	PR	06-APR-1993;	93US-004527H	
XX	PR	22-FEB-1996;	96US-060537B.	
XX	PR	23-FEB-1995;	95US-039267B.	
PA	(FORS-) FURSYTH DENTAL INFLAMRY P-R CHILDREN			
PI	Slashenko P, Li Y, Wucherplemny AL;			
DR	WP1: 2002-536031/57.			
DR	N-PSDR: ABB86734.			
PT	Novel isolated osteoclast-specific or -related DNA sequence, useful for			
PT	producing gene products useful in the therapeutic treatment or			
PT	diagnosis of disorders involving aberrant bone resorption -			
XX				

[illegible]

Db 242 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
 QY 827 GGTGTTACTATGAAATCATCTGTT---ACTGAAATGTAATCATGATATTAGTGGTT 883
 Db 262 GlyAlaTyrTyrAspIleSerTyrSerAsnSerAspAsnLeuAlaValLeuAlaVal 281
 QY 884 GGTATGTTAACTTATGCGAAACACTACTGCTTGTGCAAAAACAGCTGGCCCTCAAC 943
 Db 282 GlyTyrGlyIleGlnLysGlyAsnLysHisTrpIleLeuLysAsnSerTrpGlyLys 301
 QY 944 TTGGGACCAAGATATATTCGGATGCGAAGAACAGTGAATCAGCTGGGATTGCT 1003
 Db 302 TrpGlyAsnLysGlyTyrIleLeuMetAlaIleArgAsnLysAsnAlaIleGlyIleAla 321
 QY 1004 AGTTATCCTCTTACCCAGAAATC 1027
 Db 322 AsnLeuAlaSerPheProLysMet 329
 RESULT 13
 ID AAM48939 standard; Protein: 329 AA.
 AC AAM48939:
 DE 26-OCT-1998 (first entry)
 XX
 DE Mutant human cathepsin K C139A protease.
 XX
 KW Mutant human cathepsin K protease; SPA; tyrosine phosphatase;
 KW scintillation proximity binding assay; diabetes; cancer;
 KW 4-phosphono(difluoromethyl) phenylalanine group; osteoporosis.
 OS Homo sapiens.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..15 /note= "Pre-domain"
 FT Peptide 16..114
 FT Protein 115..329 /note= "Pro-domain"
 FT Misc-difference 139 /note= "Human cathepsin K C139A"
 FT /note= "changed from Cys in wild-type to Ala in mutant."
 PN W09820024-A1.
 PD 14-MAY-1998.
 XX
 PF 03-NOV-1997; 97WO-CA00824.
 XX
 PR 04-MAY-1996; 96US-0030411.
 XX
 PA (MERK) MERCK FROST CANADA INC.
 XX
 PI Desmarais S, Friesen R, Zamboni R;
 DR WPI: 1998-348101/30.
 DR N-PSDH: AAV32610.
 XX
 PT Peptide(s) useful in binding assays for tyrosine phosphatases or
 PT cysteine proteases - contain two or more 4-phosphono(difluoromethyl)
 PT phenylalanine groups to improve binding affinity
 XX
 PS Disclosure: Pages 43-44; 59pp; English.
 CC The present sequence represents a mutated human cathepsin K C139A
 CC protease. The invention provides a method for use in a scintillation
 CC proximity binding assay (SPA) for proteases and phosphatases. The
 CC method involves using mutated proteases and phosphatases whereby the
 CC catalytic cysteine residue of the enzymes are replaced with a serine or
 CC alanine residue to correct the problem of interference in SPA from

CC extraneous oxidizing and alkylating agents. The mutation affects the
 CC catalytic properties of the enzyme but does not affect their binding
 CC properties. The invention claims for new ligands for use in SPA which
 CC have increased binding affinity for a tyrosine phosphatase or cysteine
 CC protease. The ligands contain at least two 4-phosphono(difluoromethyl)
 CC phenylalanine groups which increase binding affinity of the ligand to
 CC its respective enzyme. The assay can be used to determine the ability
 CC of new ligands and compound mixtures to competitively bind with an
 CC enzyme. The method is claimed to allow a better usage of SPA in the
 CC discovery of compounds for the treatment and study of diseases, e.g.
 CC diabetes, cancer and osteoporosis.
 XX
 SQ Sequence 329 AA;
 Alignment Scores:
 Pred. No.: 3 35e-94 Length: 329
 Score: 930.50 Matches: 182
 Percent Similarity: 70.128 Conservative: 48
 Best Local Similarity: 55.498 Mismatches: 89
 Query Match: 46.598 Indels: 10
 DB: 19 Gaps: 4

US-10-010-580-1 (1-1126) x AAM48939 (1-329)

QY 54 CTGCTTCGCTTGTGCTGCTATGCGATGCGCAACATGATTAAGATCCACTGCGATCAT 113
 Db 8 LeuLeuProValValSerPheAla-----LeuTyrProGluGluIleLeuAspThr 24
 QY 114 CACTGGAATCTCTGGAAGAAACCTACAGCAACATCAAGAGAGAT-GAGGAGTA 172
 Db 25 HistProLeuLeuTrpLysLysThrHisArgLysGlnTyrAsnAsnLysValAspGluIle 44
 QY 173 GCACGGGCTCATCTGCGGAAAAAATCTAAATTTGTGATGCTTGCATCAATCTGCAAC 232
 Db 45 SerArgArgLeuLeuTrpGluLysAsnLeuLysTyrIleSerIleHisAsnLeuGluAla 64
 QY 233 TCAATGGCAATGCATGATATGATGATCTAGCATCAACCATCTGGAGACATGATCTGCA 292
 Db 65 SerLeuGlyValHisThrTyrGluLeuAlaMetAsnHisLeuGlyLysPheThrSerGlu 84
 QY 293 GAAGTATATCTTGTATGGGTTCCCTGAGAGTTCCACGCCATAGCAGAG-----AAT 346
 Db 85 GluValValGlnLysMetThrGlyLeuLysValProLeuSerHisSerArgSerAsnAsp 104
 QY 347 GTCACTTATAGTCAAACTGATCAAGAAATGGCTATCTGTGACTGAGAGAGAGAG 406
 Db 105 ThrLeuTyrIleProGluTrpGluGlyArgAlaIleProAspSerValAspTyrArgLys 124
 QY 407 GCGTCTGTACTGAAGTAATACAGGGTTCTTGCTGCTGCTTGGCTTTACAGCC 466
 Db 125 GlyTyrValThrProValLysAsnGlnGlyGlnCysGlySerAlaTrpAlaPheSerSer 144
 QY 467 GTGGGGGCCCTGGGAAGCAGCACTGAGCTTAAACAGAGAAAGTGGTGTCTTGAAGTGA 526
 Db 145 ValGlyAlaLeuGlnGlyLeuLeuLysLysLysThrGlyLysLeuLeuAsnLeuSerPro 164
 QY 527 CAGAACTGTAGATGCTCACTGAAATATATGGAATAAAGGCTGCAATGAGGCTTTC 586
 Db 165 GlnAsnLeuValAspCysValSerGlu-----AsnAspGlyCysGlyGlyGlyTyr 181
 QY 587 ATGACAACCTGCTTTCATATATATATATATCAACAAGCGCATCTGACAGCTTCCAT 646
 Db 182 MetTrpAsnAlaPheGlnTyrValGlnLysAsnArgGlyIleAspSerGluAspAlaTyr 201
 QY 647 CCTTACAAAGCCATGATGAGAAAGTCAATATATGATATGATGATGATGATGATGAT 706
 Db 202 ProTyrValGlyClnGlnGlnLysCysMetTyrAsnProThrGlyLysAlaIleAlaCys 221
 QY 707 TCAAACTATGAACTTCCCTTTGGCAGTGAAGATGCTTAAAGATGATGATGAT 766
 Db 222 ArgGlyTyrArgGluIleProGluGlnLysGlnLysAlaLeuLysAlaValAlaArg 241
 QY 767 AAAGACCTGTGTGTGTGCTATAGATGAGAGCCACTATCTTCTTCTCTTACAGAGT 826


```
UY 767 AAAGACCTGTGCTGCTATAGATGCCAGCCACTATCTTCTCTTACAGAACT 826
    |||
DB 242 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
    |||
UY 827 GATGCTTACATATGACCAATCCCTGT--ACTGAGAAATGTGAATGATGAGATATATAGTGGTT 883
    |||
DB 262 GlyValTyrTyrAspGluSerCysAsnSerAspAsnLeuAsnHisAlaValIleGluAlaVal 281
    |||
UY 884 GACTATGGTACCTTAATGGCAAGATCTGCTGTGTGAAAATATATGTTGGCTGAC 943
    |||
DB 282 GlyTyrGlyIleGlnIleGlyAsnIleHisTyrPheIleIleLysAsnSerTTPGlyGluAsn 301
    |||
UY 944 TTGCTGACCAAGCATATATTCGATGCCAAGAAACAGTGCAGATCTGTGCATTGCT 1003
    |||
DB 102 TrpGlyAsnLysGlyTyrIleLeuMetAlaIleGluAsnLysAsnAsnAlaCysGlyIleAla 321
    |||
UY 1004 ATTATGCTGCTCTTACCCAGAAATC 1027
    |||
DB 122 AsnLeuAlaSerPheProLysMet 329
    |||
```

Search completed: June 27, 2003, 11:28:42
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:23:59 ; Search time 49.5 Seconds
(without alignments)
4373.631 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1997
Sequence: 1 gcattatgattatgagagca.....aataatgtcatgaccat 1126

Scoring table: BIOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96334422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US10010580/unat_27062003.104536.11748/app_query.fasta.1.1287
-DB=PIR.73 -QFMT=fastan SUPP1X=n2p.rpr MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010580.CGN.1.1.62 @runat.27062003.104536.11748 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESOURCE -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR.73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1562	78.2	331 2	A42482 cathepsin S (EC 3. cathepsin S (EC 3. cathepsin S (EC 3.
2	1334.5	66.8	330 2	A45087 cathepsin S (EC 3. cathepsin S (EC 3. cathepsin S (EC 3.
3	1089	54.5	217 2	S15884 cathepsin K (EC 3. cathepsin K (EC 3. cathepsin K (EC 3.
4	944.5	47.3	329 2	A49868 cathepsin K (EC 3. cathepsin K (EC 3. cathepsin K (EC 3.
5	939.5	46.0	329 2	JC2476 cathepsin K (EC 3. cathepsin K (EC 3. cathepsin K (EC 3.
6	922.5	47.2	329 2	S74227 cathepsin K (EC 3. cathepsin K (EC 3. cathepsin K (EC 3.
7	835.5	41.8	334 1	KHMSL cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
8	822.5	41.2	334 1	KHMSL cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
9	806	40.4	333 1	KHHUL cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
10	804	40.3	334 2	A58195 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
11	781	39.1	313 2	S47433 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
12	777	38.9	323 2	S19650 cysteine proteinases cathepsin L-like c
13	770.5	38.6	338 2	JC5443 cysteine proteinases cathepsin L-like c
14	768.5	38.5	337 2	T24387 probable cysteine

15	764	38.3	326 2	S53027 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
16	760.5	38.1	339 2	A53810 cysteine endopepti cysteine endopepti cysteine endopepti
17	760.5	38.1	344 2	JX0366 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
18	757	37.9	320 2	S19651 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
19	748	37.5	333 1	I52525 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
20	734	36.8	343 2	JC7183 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
21	727.5	36.4	331 2	JC5441 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
22	724.5	36.3	324 2	JC5442 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
23	724.5	36.3	324 2	JC5442 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
24	715.5	35.8	322 2	S19649 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
25	709.5	35.5	326 2	S43991 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
26	706.5	35.4	480 2	T01207 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
27	697.5	34.9	218 1	KHCHL cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
28	684.5	34.3	218 2	S67481 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
29	675	33.8	458 1	KHRZO cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
30	667	33.4	452 2	JN0719 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
31	660.5	33.1	326 2	T09259 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
32	654	32.7	368 2	S47312 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
33	651.5	32.6	317 2	S44151 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
34	647	32.4	471 1	KHRZO cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
35	644	32.2	454 2	JC4848 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
36	639.5	32.0	357 2	T06529 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
37	638	31.9	455 2	T12041 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
38	637	31.9	453 2	T01206 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
39	631	31.6	427 2	S57776 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
40	623.5	31.2	364 2	T12039 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
41	621.5	31.1	466 2	T06416 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
42	620	31.0	331 2	D86413 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
43	619	31.0	376 2	E85435 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
44	618.5	31.0	364 2	T46630 cathepsin L (EC 3. cathepsin L (EC 3. cathepsin L (EC 3.
45	617.5	30.9	376 1	KHDOF prestatik cathepsin

ALIGNMENTS

RESULT 1
A42482
cathepsin S (EC 3.4.22.27) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Mar-1993 #sequence, revision 18-Nov-1994 #ext_change 22-Jun-1999
C:Accession: A42482; A53625; A42896
R:Shi, G.P.; Munger, J.S.; Weira, J.P.; Rich, D.H.; Chapman, H.A.
J. Biol. Chem. 267, 7258-7262, 1992
A:Title: Molecular cloning and expression of human alveolar macrophage cathepsin S, a
A:Reference number: A42482; M0ID:92218373; PMID:1377692
A:Accession: A42482
A:Molecule type: mRNA
A:Residues: 1-331 <SH1>
A:Cross-references: GB:S9414; MID:9248405; PIDN:AAB22005.1; PID:9248406
A:Experimental source: alveolar macrophage
A:Note: sequence extracted from NCBI backbone (NCBIN:93414, NCBIIP:93443)
P:Shi, G.P.; Webb, A.C.; Foster, K.E.; Knoll, J.H.M.; Lemere, C.A.; Munger, J.S.; Cha
J. Biol. Chem. 269, 11530-11536, 1994
A:Title: Human cathepsin S: chromosomal localization, gene structure, and tissue dist
A:Reference number: A53625; M0ID:94209337; PMID:8157683
A:Accession: A53625
A:Molecule type: DNA
A:Residues: 1-210, 'H', 212-331 <SH2>
A:Cross-references: GB:U07374
R:Wiederanders, B.; Brumme, D.; Kirschke, H.; von Figura, K.; Schmidt, B.; Peters, C.
J. Biol. Chem. 267, 13708-13713, 1992
A:Title: Phylogenetic conservation of cysteine proteinases. Cloning and expression of
A:Reference number: A42896; M0ID:92317106; PMID:1377692
A:Accession: A42896
A>Status: not compared with conceptual translation
A:Molecule type: mRNA; protein
A:Residues: 1-91, 'W', 93-160, 'S', 162-331 <WIE>
A:Cross-references: GB:S19127; GB:M90696; MID:9250802; PIDN:AAA09269.1; PID:9250803
A:Experimental source: testis
A:Note: sequence extracted from NCBI backbone (NCBIIP:107806)
C:Genetics:
A:Gene: GDB:GTSS
A:Cross-references: GDB:132414; OMIM:116845

A: Map position: 1q21-1q21

C: Superfamily: papain

C: Keywords: cysteine protease; hydrolase; lysosome

F: 1-16/Domain: signal sequence #status predicted <SIG>

F: 17-114/Domain: propeptide #status predicted <PRO>

F: 115-331/Product: cathepsin S #status predicted <CAT>

F: 139,278,298/Active site: Cys, His, Asn #status predicted

Alignment Scores:

pred. No.:	4,98e-128	length:	331
Score:	1562.00	Matches:	288
Percent Similarity:	93.05%	Conservative:	20
Best Local Similarity:	87.01%	Mismatches:	23
Query Match:	78.22%	Indels:	1
DB:	2	Gaps:	0

US-10-010-580-1 (1-1126) x A42482 (1-331)

```
QY 36 ATGAATGCTTATGTTGGCTTCTGTTGCTGCTATGCGAGTGGCAAGTACATMAA 95
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
QY 96 GATGCCAGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 AsproThrLeuAspHisHisThrHisLeuThrLysLysThrLysGlnLys 40
QY 155 GAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAG 214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 GluLysAsnGlnGlnLysAlaValAlaArgArgLeuLeuTrpLysAsnLeuLysPheValMet 60
QY 215 CTTTCAATATCTGAAACACTCAATGGAATGATGATGATGATGATGATGATGATGATG 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerLysAspLeuGlyMetAsnHisLeu 80
QY 275 GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 81 GluAspMetThrSerGlnGlnValMetSerLeuThrSerSerLeuAlaGlnValProSerGln 100
QY 335 TGCGACAGAAATGTCACCTTATAGTCAAACTGATGATGATGATGATGATGATGATGATG 394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 TrpLysArgAsnHisThrLysSerAsnProAsnArgHisLeuProAspSerValAsp 120
QY 395 TGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TrpArgGlnLysGlnLysValThrGlnValLysTrpGlnLysSerGlnLysAlaCysTrp 140
QY 455 GCTTTGACGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTG 514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 141 AlaPheSerAlaValGlnLysAlaLeuGlnAlaGlnLeuLysLeuLysThrGlnLysVal 160
QY 515 TCTCTGACGTGACGAGAACTTGTGATGATGATGATGATGATGATGATGATGATGATG 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 161 ThrLeuSerAlaGlnHisLeuValAspCysSerThrGlnLysTrpGlnLysGlnLys 180
QY 575 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 AsnGlnGlnPheMetThrThrAlaPheGlnTrpHisLeuAspAsnLysGlnLysLeuAspSer 200
QY 635 GAACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 694
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 AspLysSerLysProGlnLysAlaMetAlaGlnLysGlnLysGlnLysGlnLysGlnLys 220
QY 695 GCTGCCAGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 754
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 AlaAlaThrCysSerLysTrpThrGlnLeuProGlnLysTrpGlnLysValLeuLysGln 240
QY 755 GCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AlaValAlaAsnLysGlnLysValSerValGlnLysAlaPheAlaGlnHisProSerPhePhe 260
QY 815 CTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 LeuLysArgSerGlnLysValLysTrpGlnProSerCysThrGlnAsnValAsnHisGlnVal 280
```

```
QY 875 TTAGTGGTGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 934
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 281 LeuValValGlnLysTrpLysLeuAsnGlnLysGlnLysTrpLeuValLysAsnSerTrp 300
QY 935 GGCCTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 994
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GlyHisAsnPheGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 320
QY 995 GGCATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 GlyIleAlaSerPheProSerTrpProGlnLys 331
```

RESULT 2

A45087

cathepsin S (EC 3.4.22.27) - rat

C: Species: Rattus norvegicus (Norway rat)

C: Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 22-Jun-1999

C: Accession: A45087

R: Petanceska, S.; Devi, L.

J. Biol. Chem. 267, 26038-26043, 1992

A: Title: Sequence analysis, tissue distribution, and expression of rat cathepsin S.

A: Reference number: A45087; M01D:93100327; PMID:1281481

A: Accession: A45087

A: Status: preliminary

A: Molecule type: mRNA

A: Residues: 1-330 <PEP>

A: Cross-references: GB:103201; NID:q203649; PIDN:AAA0994.1; PID:q203650

A: Experimental source: brain

A: Note: sequence extracted from NCBI backbone (NCBIN:120879, NCBI:120880)

C: Superfamily: papain

C: Keywords: cysteine protease; glycoprotein; heterodimer; hydrolase; lysosome

F: 137,277,297/Active site: Cys, His, Asn #status predicted

Alignment Scores:

pred. No.:	3,38e-108	length:	330
Score:	1334.50 <td>Matches:</td> <td>257</td>	Matches:	257
Percent Similarity:	85.54% <td>Conservative:</td> <td>27</td>	Conservative:	27
Best Local Similarity:	77.41% <td>Mismatches:</td> <td>44</td>	Mismatches:	44
Query Match:	66.83% <td>Indels:</td> <td>5</td>	Indels:	5
DB:	2 <td>Gaps:</td> <td>3</td>	Gaps:	3

US-10-010-580-1 (1-1126) x A45087 (1-330)

```
QY 41 ATGCGTAGTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MetAlaValLeuGlnLysAlaProGlnLysCysAspAsnGlnLysAlaThrAlaLysArg-Pr 20
QY 101 CACTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20 ThrLeuAspHisHisThrAspLeuTrpLysLysThrArgMetArgAsnThrAspGln 40
QY 160 GAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 AsnGlnGlnLysAspValAlaArgLeuLeuLeuTrpLysAsnLeuLysPheIleMetLeuHis 60
QY 220 CAATCTGGAACACTCAATGGAATGATGATGATGATGATGATGATGATGATGATGATG 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 AsnLeuGlnHisSerMetGlyMetHisSerLysSerValGlnMetAsnHisMetGlnLys 80
QY 280 CATGATGCTGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 PheThrProGlnGlnValLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 340 GAGAAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 AsnSerGlnLysLeuLysSerSerSerSerSerSerSerSerSerSerSerSerSerSer 120
QY 400 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 GlnLysGlnLysValThrAsnValLysTrpGlnLysSerCysGlnLysSerCysTrpAlaPhe 140
QY 460 CAGCGGTGCGGGGCGGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
```


QY 927 ACACCTGGGCGCCACACTTCGTCACCAAGCATATATTCGACATCCAGCAAGACAGTGGAA 986
|||||
Db 300 snu5ttrpGlyIyLgltItrpGlyMetAspGlyTrpIleIleAlaIyAspArgAsnA 320
|||||
QY 987 ATACGTCTGGATTCCTACCTTATCCCTCTTACCCA 1021
|||||
Db 320 snu5ttrpGlyIyLgltItrpGlyMetAspGlyTrpIleIleAlaIyAspArgAsnA 331
|||||

RESULT 9

KIHHL

N:Alternate names: precursor (validated) - human
C:Species: Homo sapiens (man)

C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 08-Dec-2000
C:Accession: S01002; B32333; S09065; A45043; S00323; B27011; A26069; A32683; D27011; F27

R:Gal, S.; Gottesman, M.M.
Biochem. J. 253, 303-306, 1988

A:Title: Isolation and sequence of a cDNA for human pro-(cathepsin L).
A:Reference number: S01002; MUID:88339905; PMID:3421948

A:Accession: S01002
A:Molecule type: mRNA

A:Residues: 1-333 <GAD>
A:Cross-references: GH:X12451; NID:g29714; PIDN:CAA30981.1; PID:g29715

R:Joseph, L.; Chang, L.C.; Stamenkovich, D.; Sukhame, V.P.
J. Clin. Invest. 81, 1621-1629, 1988

A:Title: Complete nucleotide and deduced amino acid sequences of human and murine prepro
A:Reference number: A92768; MUID:88213715; PMID:2835338

A:Accession: B32333
A:Molecule type: mRNA

A:Residues: 1-333 <JOS>
A:Cross-references: GH:M20496; NID:g809235; PIDN:AAA66974.1; PID:g190418

R:Joseph, L.; Lapid, S.; Sukhame, V.
Nucleic Acids Res. 15, 3186, 1987

A:Title: The major rat induced protein in NIH3T3 cells is cathepsin L.
A:Reference number: S09065; MUID:87174843; PMID:3550705

A:Accession: S09065
A:Molecule type: mRNA

A:Residues: 11-154 <JOS>
A:Cross-references: EMBL:X05256; NID:g29718; PIDN:CAA28877.1; PID:g1340178

R:Chauban, S.S.; Popescu, N.C.; Ray, D.; Fleischmann, R.; Gottesman, M.M.; Troen, B.R.
J. Biol. Chem. 268, 1039-1045, 1993

A:Title: Cloning, genomic organization, and chromosomal localization of human cathepsin
A:Reference number: A45043; MUID:9123212; PMID:8419312

A:Accession: A45043
A:Molecule type: DNA

A:Residues: 40-46;82-86;130-135;205-210;259-264;299-304 <CHA>
A:Cross-references: GH:I06426

A:Note: only exon-intron splice junctions are shown
R:Ritonja, A.; Popovic, T.; Kotnik, M.; Machleidt, W.; Turk, V.
FEBS Lett. 228, 441-445, 1988

A:Title: Amino acid sequences of the human kidney cathepsins H and L.
A:Reference number: S00322; MUID:88137635; PMID:3342889

A:Accession: S00322
A:Molecule type: Protein

A:Residues: 'X',115-129,'M',131-133,'E',135-141;292-307,'TD',310-333 <RTT>
R:Machleidt, W.; Ritonja, A.; Popovic, T.; Kotnik, M.; Brzin, V.; Machleidt, T.
In Cysteine Proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter,
A:Title: Human cathepsins B, H and L: characterization by amino acid sequences and some

A:Reference number: A27011
A:Accession: A27011
A:Molecule type: Protein

A:Residues: 'X',115-129,'M',131-133,'E',135-141;292-307,'TD',310-333 <MA2>
R:Mason, R.W.; Walker, J.E.; Northrop, F.D.
Biochem. J. 240, 373-377, 1986

A:Title: The N terminal amino acid sequences of the heavy and light chains of human cathepsin
A:Reference number: A26069; MUID:87127952; PMID:3545185

A:Accession: A26069
A:Molecule type: Protein

A:Residues: 114-147,'P',149-152,'Y',292-333 <MAS>
R:Smith, S.M.; Gottesman, M.M.
J. Biol. Chem. 264, 20487-20495, 1989

A:Title: Activity and deletion analysis of recombinant human cathepsin L expressed in Es

A:Reference number: A32683; MUID:90062183; PMID:2684978

A:Contents: annotation
C:Genetics:

A:Gene: GDB:CTSL
A:Cross-references: GDB:119824, OMIM:116880

A:Map position: 9q22.1-9q22.2
A:Introns: 42/3; 83/3; 132/3; 207/3; 262/1; 301/2

C:Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C:Function:

A:Description: catalyzes hydrolysis of peptide bonds in proteins
A:Pathway: Intracellular protein degradation

A:Note: Important role in the lysosomal degradation of proteins
C:Superfamily: papain

C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; prot
F:1-17/Domain: signal sequence *status predicted <STO>
F:118-113/Domain: propeptide *status predicted <PRO>

F:114-333/Product: cathepsin L *status experimental <MAT>
F:114-288/Product: cathepsin L heavy chain *status experimental <HCH>

F:292-333/Product: cathepsin L light chain *status experimental <LCH>
F:135-178,169-211,269-322/Disulfide bonds: *status predicted
F:138,276,300/Active site: Cys, His, Asn *status predicted

F:221/Binding site: carbonylate (Asn) (covalent) *status predicted

Alignment Scores:

Pred. No.:	3,986-62	Length:	333
Score:	806.00	Matches:	163
Percent Similarity:	64.74%	Conservative:	50
Best Local Similarity:	49.54%	Mismatches:	96
Query Match:	40.36%	Indels:	20
DB:	1	Gaps:	7

US-10-010-580-1 (1-1126) x KIHHL (1-333)

QY	75	GCAGTGGCAACAGTATCAATGATCCATCTGATCATCACTGGAATCTTGAGAGAA	134
Db	15	AlaSerAlaThrLeuThrPheAspSerLeuGluAlaGlnTrpThrLysTrpLysAla	34
QY	135	ACCTACGCAACAAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	194
Db	35	MetHisAsnArg-LeuTrpGlyMetAsnGluGluGlyTrpArgArgAlaValTrpGlu	54
QY	195	AAATGTAAATTTTGATGCTTCACATCTGCAACATCGCAACATCGCAACATCG	254
Db	54	YAsnMetLysMetIleGluLeuHisAsnGlnIleuTrpArgGluGlyLysHisSerPhe	74
QY	255	ATCTAGCATGACCATCTGGAGACATGACTGTGAAGAACTATATCTTGATGGTT	314
Db	74	hMetAlaMetAsnAlaPheGlyAspMetThrSerGlnGlnPheArgGlnValMetAsn	94
QY	315	CCCTG-----ACAGTCCGACGCA-----TGCCAGAGAAATGCTACTTAACT	359
Db	94	LypPheGlnAsnArgLysProArgLysGlyLysValPheGlnIleuProLeuPheTrpGlu	114
QY	360	CAACATCTATCATCAAAATGGCTGATCTGCTGAGTGGAGAGAGAGAGAGAGAG	419
Db	114	La-----ProArgSerValAspTrpArgGluLysGlyValThrP	128
QY	420	AAATGAAATACCAAGGTTTGTGTGCTGTGTTGAAATTTAAAGTGTGAGGCTGG	479
Db	128	roValLysAsnGlnGlyLysCysGlySerCysTrpAlaPheSerValThrGlyAlaLeu	148
QY	480	AACACACACTGAAAGCTAAACAGCAAGAGCTGCTCTGCTGAGTGGAGAGAGAG	539
Db	148	LgylGlnMetPheArgLysThrArgLysGlnLeuSerLeuSerLeuGlnLeuValA	168
QY	540	ATTGCTCAACTGAAATATGGAATTAAGCTGCAATGCGCTTTCATGCAACCTGCT	599
Db	168	spCysSerLysProGln---GlyAsnGlnGlyCysAsnGlyCysLeuMetAspTrpAla	187
QY	600	TCCAGTATATTTTATATTAACAGAGCGATGATGAGAGCTTCATCTCATCAAGCA	659
Db	187	hGlnTrpValGlnAsnGlnGlyLysLeuAspSerGluSerGluSerTrpProTrpGlu	207

OY 660 TGAATGCAAGTGCAGATATGACTCAAAAAGCCAGCTCCACATGTCGAAAGTATCTG 719
 Db 207 hrGluGluSerCysLysTyrAsnProLysSerValAlaAsnAspThrGlyPheValA 227
 OY 720 AACTCCCTTTGGCAGTGAAGATGCCCTTAAGAACAGCTGGCCATTAAGACCTGCT 779
 Db 227 spLlePro---LysGlnGluLysAlaLeuMetLysAlaValAlaThrValGlyProIles 246
 OY 780 CTGTGCTATAGATGCGAGCCACTATCTCTTCTCTGACAGAGTGGTGTCTACTATG 839
 Db 246 erValAlaIleAspAlaGlyHisGlnSerPheLeuPheLysGlnGlyLysIleTyrPheG 266
 OY 840 AACCATCTCT---ACTCAGCAATGTGATATGAGCATATGATGCTGCTGCTATGCT 892
 Db 266 LurProAspCysSerSerGluAspMetLysPheLysGlyValLeuValValGlyTyrGlyPheG 286
 OY 893 -----AACCTTAAGGCAAGACACTAGCTGCTTGTGAAAAACAGCTGGGCTCACT 944
 Db 286 LurSerThrGluSerAspAsnAsnLysTyrThrPheValLysAsnSerThrPylGluGlu 306
 OY 945 TTGGTGACCAAGATATATTCGATGCGAGAAACAGTGAATCACTGTGGATTCCTA 1004
 Db 306 rpgLysMetGlyGlyTyrValLysMetAlaLysAspArgAsnHisCysGlyIleAla 326
 OY 1005 GTTATCCCTTACCCCAAGAAATC 1027
 Db 326 erAlaAlaSerTyrProThrVal 333

RESULT 10

A58195

N:Alternates: L (EC 3.4.22.15) precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 16-Oct-1996 #sequence_revision 18-Oct-1996 #text_change 22-Jun-1999

C:Accession: A58195, S59911, S59916

R:Okamura, N.; Tamura, M.; Uchiyama, Y.; Sugita, Y.; Daehue, F.; Syntin, P.; Daehue, J.

Biochim. Biophys. Acta 1245, 221-226, 1995

A>Title: Direct evidence for the elevated synthesis and secretion of procathepsin L in t

A:Reference number: S59911; MID:96069841; PMID:7492581

A:Accession: A58195

A:Molecule type: mRNA

A:Residues: 1-334 <OKA>

A:Cross-References: DDBJ:D37917; NID:q710655; PIDN:BA07140.1; PID:q1468964

A:Experimental source: epididymis

A>Note: The authors translated the codon CCG for residue 203 as Thr and CCC for residue

A>Note: submitted to the DDBJ/EMBL/Genbank databases by Naomichi Okamura, 8 August 1994

A:Accession: S59911

A:Molecule type: protein

A:Residues: 18-31 <OK2>

C:Complex: heterodimer of disulfide linked chains produced from a single chain precursor

C:Function:

A:Description: catalyzes hydrolysis of peptide bonds in proteins

A:Pathway: protein degradation

A>Note: important role in the lysosomal degradation of proteins

C:Superfamily: papain

C:Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein

F:1-17/Domain: signal sequence #status predicted <STIC>

F:18-113/Domain: propeptide #status predicted <PRO>

F:114-334/Product: cathepsin L #status predicted <MAT>

F:135-178/169-212/270-323/Disulfide bonds: #status predicted

F:138,277,301/Active site: Cys, His, Asn #status predicted

F:222,292/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	5,94e-62	Length:	334
Score:	804.00	Matches:	164
Percent Similarity:	63.72%	Conservative:	52
Best local Similarity:	48.38%	Mismatches:	111
Query Match:	40.26%	Indels:	12
DB:	2	Gaps:	7

US-10-010-580-1 (1-1126) x A58195 (1-334)

OY 36 ATGAATGCTAGTGGCGCTGCTCCGTTGTC---TCTATGCAAGTGCACAGACATGAT 92
 Db 1 MelLysProSerLeuPheLeuThrAlaLeuCysLeuGlyIleAlaSerAlaIleProLys 20
 OY 93 AAGATCCCACTGCTGATCATCATGGAATCTCTGGAAGAAAACCTACAGCAACATCA 152
 Db 21 LeuAspGlnAsnLeuAspAlaAspTrpTyrLysTrpLysAlaIleHisGlyArg---LeuT 40
 OY 153 AGGAAGCAATGAGGAAGTGCACAGCGGCTGCTCATCTGGGAAAAAATCAAAATTGTGA 212
 Db 40 yrgLysMetLysGlnGlyLysTrpArgArgAlaValIleTrpLysAsnMetLysMetIleG 60
 OY 213 TGCTTCAATCTGGAACACTCAATGGCAATGCATTCATATGATCAGGATGACCAATC 272
 Db 60 LuleHisAsnGlnGlnLysSerLysLysHisGlyPheSerMetAlaMetLysAlaP 80
 OY 273 TGGGAGCATGACTGCTGGAAGAGTGTATCTTGTAGGTTCCCTGAGAGTTCCAGGC 332
 Db 80 heGlyAspMetThrAsnGlnGlnLysPheArgGlnValMetAsnGlyPheGln---AsnGln 99
 OY 333 AATGCAAGCAAAATGTCACCTATAGCTCAAACTCTAATCAGAAATGCTGATTCGTG 392
 Db 99 yshLysLysGlyLysValPheHisGlnSerLeuValLeuGluValProLysSerValA 119
 OY 393 ACTGAGAGAGAGAGCGCTGCTTACTGAAATGCAATACAGGCTTCTGCTGCTGCTT 452
 Db 119 sPTPrArgGlnLysGlyLysValThrAlaValLysAsnGlnGlnLysGlySerCys 139
 OY 453 GGGCTTCAAGCGCTGCTGCTGCTGGAAGCAAGCAAGTGAATGAAGCAAGAGCTG 512
 Db 139 rPalPheSerAlaThrGlyLysLeuGlnGlnMetPheArgLysThrGlyLysLeu 159
 OY 513 TGCTCTGAGTGCACAGAACTTGTGATGCTCTCACTGAAATATGGAATTAAGGCT 572
 Db 159 alSerLeuSerGlnGlnLysLeuValAspCysSerArgProGln---GlyAsnGlnGly 178
 OY 573 GCAATGGCGGCTTCACTGCAACTGCTTCCAGTATATATGATACACAGCGCATTCAT 632
 Db 178 yAsnGlnGlyLeuMetLysAsnAlaPheGlnTyrValLysAspAsnGlyLysLeuAsp 198
 OY 633 CAGAACTCTATCTCCCTACAAAGCCATG---ATGGAATGCGAATATGACCAAAA 689
 Db 198 hrGlnGlnSerTyrProTyrLeuGlnLysArgGlnThrAsnSerCysThrTyrLysProGlu 218
 OY 690 AGGAGCTGCGACATGTTCAAAATGATCTGAACTCCCTTGGCAGTGAAGATGCTTAA 749
 Db 218 ySerAlaAlaAsnAspThrGlyPheValAspIlePro---GlnArgLysAlaLeu 237
 OY 750 AAGCACTGTGCGCAATAAAGAGCTGTGCTGCTGCTATAGATGCGACGACTATTC 809
 Db 237 erLysAlaValAlaThrValGlyProIleSerValAlaIleAspAlaGlyHisSerSer 257
 OY 810 TCTCTCGTACAGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
 Db 257 heLysPheTyrLysSerGlyLysIleTyrLysProAspCysSerSerLysAspLeuAsp 277
 OY 867 ATGAGTATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914
 Db 277 lsgLysValLeuValAlaGlyTyrGlyPheGlnGlyThrAspSerAsnSerSerLysPhe 297
 OY 915 GCGCTTGAAGAAACAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974
 Db 297 rLysValLysAsnSerTrpGlyProGlnGlnLysTrpAsnGlyLysValLysMetAla 317
 OY 975 GAAACAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
 Db 317 yAspArgLysAsnAsnHisCysGlyIleSerThrAlaAlaSerTyrProThrVal 334

RESULT 11

S47433

N:Alternates: L (EC 3.4.22.15) - Norway lobster

C:Species: Nephrops norvegicus (Norway lobster)

C>Date: 23-Nov-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999


```

OY 287 GGTGAAGAGATGATATCTTTCATGCGTTCCTGAGAGTTCGCCACCAATGGCAGAGA--- 343
    |||||
Db 77 LeuGIuGluPheAsnAlaValMetLysGly---AsnIleProArgSerAlaProVal 95
OY 344 AATGCACTTATAGGTCAAACTCAATCACAATATGGCTGATCTTGCGATCGAGAGAG 403
    |||||
Db 96 SerValIlePheTyrProLysLysGIuThrGIuProGlnAlaThrGIuValAspTrrArgThr 115
OY 404 AAGGCGTGTGTACTGAAGTAAATACCAGGGTCTTGCTGCTTGTGGCTTTACG 463
    |||||
Db 116 LysGIuAlaValAlaThrProValLysAspGlnGlyLysGIuSerCysTrrPalaPheSer 135
OY 464 GCTGCGGGCGCCGCGAAGCACAACGTGAAGCTAAAAACAGAAAGCGTGTCTGAGT 523
    |||||
Db 136 ThrThrGlySerLeuGIuGlyGlnHisPheLeuLysThrGlySerLeuIleSerLeuAla 155
OY 524 GCACAGAACTTGATAGATTGCTCAACATAAATAATGGGAATAAAGCGTCGAATGGCGT 583
    |||||
Db 156 GluGlnGlnLeuValAspCysSer---ArgProTyrGIuProGlnGlyCysAsnGIuYol 174
OY 584 TTCATGACAACCTGCTTCCATATATATATGATAAAGCGCATTTGATCGAAGCTTCC 643
    |||||
Db 175 TrpMetAsnAspAlaPheAspTrrIleLysAlaAsnAsnGIuIleAspThrGIuAlaAla 194
OY 644 TATCCCTACAAAGCCCATGAAATGGAAGTGCAGATATGACTCAAAAAACCGAGCTGCCACA 703
    |||||
Db 195 TyrProTyrGIuAlaAlaArgAspGIuSerCysArgPheAspSerAsnSerValAlaAlaThr 214
OY 704 TGTTCAAAGTATACGAACTTCCTTTGGCAGTGAAGTGCCTTAAAGAGCTGTGCC 763
    |||||
Db 215 CysSerGIuLysIleThrAsnIleAlaSerGIuSerGIuThrGIuLysGlnGlnAlaValArg 234
OY 764 AATAAGAGACGCTGTCTGTGCTATAGATGGAGCCACTATTTCTTCTCTGACAGA 823
    |||||
Db 235 AspIleGIuLysProLysSerValThrIleAspAlaAlaHisSerSerPheGlnPheTyrSer 254
OY 824 AGTGTGTCTACTATGAACCATCCCTGTACTCAGAAT---GTCAATCATGCACTATTAGT 880
    |||||
Db 255 SerGIuValIleTyrGIuLysProSerCysSerProSerTyrLeuAspHisAlaValLeuAla 274
OY 881 GTTGCTATGCTTAACCTTAATGGAAGTACTGCTGTGCAAAACAGCTGGCGCTGC 940
    |||||
Db 275 ValGIuTyrGIuSerGIuGIuGIuLysAspPheTrrIleValLysAsnSerTrrPalaThr 294
OY 941 AACTTGTGTGACCAAGATATATTCGATGGCAGAGAAACAGTGAATCACTGTGGATT 1000
    |||||
Db 295 SerTrrPalaAspAlaGIuTyrIleLysMetSerArgAsnArgAsnAsnAsnCysGIuIle 314
OY 1001 GCTAGTTATCCCTCTTACCCA 1021
    |||||
Db 315 AlaThrValAlaIleSerTrrPro 321

```

RESULT 13

JC5443 cathepsin L-like cysteine proteinase (EC 3.4.22.-) cl [similarity] - Maize weevil

C:Species: Sitophilus zeamais (maize weevil)

C:Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000

C:Accession: JC5443

R:Matsumoto, I.; Emori, Y.; Abe, K.; Arai, S.

J. Biochem. 121, 464-476, 1997

A:Title: Characterization of a gene family encoding cysteine proteinases of Sitophilus z

nd germ cells.

A:Reference number: JC5441: M01D:97279098; P01D:9133615

A:Accession: JC5443

A:Molecule type: mRNA

A:Residue: 1-338 <MA>

A:Cross-references: GB:082884; NID:92804261; PIDN:BAA24442.1; PID:92804262

C:Comment: This enzyme acts as a secreted or lysosomal proteinase. It has various functi

y organs, and cleaning in the malpighian tubule.

C:Superfamily: papain

C:Keywords: cysteine proteinase; hydrolase

F:145,284,305/Active site: Cys, His, Asn *status predicted

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
4,959-59	770.50	63.25%	50.30%	38.58%	2
Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
338	167	43	110	13	5

US-10-010-580-1 (1-1126) x JC5443 (1-338)

```

OY 56 GCTTCGCTTGTGCTCCATGACAGTGCACAGTACATAAAGATCCAGTCTGCATGATCA 115
    |||||
Db 8 AlaAlaValAlaIleSerCysGlnAlaValSer-PheTyrAsp---LeuValGIuGIuGI 26
OY 116 CTGGAATCTGTGGAGAAAGAAACCTACAGCAACCAATC-AAGGAAGAGATGAGCAATAC 174
    |||||
Db 26 nrrPserSerPheLysMetGIuHisSerLysAsnTyrAspSerGIuThrGIuGIuArgH 46
OY 175 ACGCGCTGCTGATGCGAAGAAATCTAAATTTGTATGCTTGCATCTGCATGCAACATC 234
    |||||
Db 46 eArgMetLysIlePheMetGIuAsnAlaHisLysValAlaLysHisAsnLysLeuPheSe 66
OY 235 AATGGCAATGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294
    |||||
Db 66 rGlnGIuPheValLysPheLysLeuGIuLysAsnLysTyrAlaAspMetLeuHisGI 86
OY 295 AGTGAATCTTGTGATG-----GGTCCCTGAGAGATTCGACGCA 333
    |||||
Db 86 urPheValSerThrLeuAsnGIuPheAsnLysThrLysAsnHisIleLeuLysGIuSerAs 106
OY 334 ATGGCAGAGAAATGCTACTATATAGCTCAAACTCAATCAAGAAATGGCTGATCTGTGCA 393
    |||||
Db 106 rLeuAsnAspAlaValAlaArgPheIleSerProAlaAsnValLysLeuProAspThrValAs 126
OY 394 CTGAGAGAGAGAGAGCGCTGTGTTACTGAACTGAAGTGAATACAGGTTCTGTGCTGCTGTG 453
    |||||
Db 126 pTrrArgAspLysGIuAlaValAlaThrGIuValLysAspGIuLysCysGIuSerCysTrr 146
OY 454 GGTCTTACGCGCTGTGGGGCGCTGGAAG-AACACTAAAGCTAAAGAAACAGAAAGCTGGT 513
    |||||
Db 146 pSerPheSerAlaThrGIuSerLeuGIuGIuGIuHisPheArgLysThrGIuLysLeuVal 166
OY 514 GTCTGCTAGTGCACAGAACTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 573
    |||||
Db 166 lSerLeuSerGIuGIuAsnLeuValAspCysSer---GIuArgTyrGIuLysAsnGIuY 185
OY 574 CAATGCGCGTTCATGCAACACTGCTTCCAGTATATATATGATTAACAACGCGATGATTC 633
    |||||
Db 185 sAsnGIuYolLysMetAspAsnAlaPheArgTyrIleLysAspAsnGIuGIuIleAspThr 205
OY 634 AAGAGCTTCCATGCTTACCAAGAGCATGATGGAAGATGCAAGATATGATGCAAAAAAGCG 693
    |||||
Db 205 rGIuLysSerTyrProTyrLeuAlaGIuAspGIuLysCysHisTyrLysAlaGIuAsnSe 225
OY 694 AGTGCACACATGTTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 753
    |||||
Db 225 rGIuAlaThrAspLysGIuPheValAspIleGIuGIuAlaAsnGIuLysAspLeuLysAl 245
OY 754 AGCTGTGCGCCATTAAGAGAGCTGTGCTGTGCTGATGATGATGATGATGATGATGATGAT 813
    |||||
Db 245 dAlaValAlaThrValGIuProValSerIleAlaAlaAspAlaSerHisGIuThrPheGI 265
OY 814 CCTGATGAGAGAGTGGTGTCTACTGTGAACATATCTGT---ACTTACATATGATATATGAG 870
    |||||
Db 265 nLeuTyrSerAspGIuValIleTyrSerAspProGIuLysSerSerGIuGIuLysAsnHisGI 285
OY 871 AGTATTACTGTGCTGCTTATGCTTAACCTTT---AATGGCAAGAGCTTACTGCTTGAAGAA 927
    |||||
Db 285 yValLeuValAlaGIuTyrGIuThrSerAspAspGIuGIuAspTrrPheValLysAs 305
OY 928 CAGCGCGCGCTCAACTTGTGCTGACCAAGATATATGCGATGCGCAAGAAACACTGCAAA 987
    |||||

```




GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:06:33 ; Search time 24 Seconds
(without alignments)
3891.862 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1997
Sequence: 1 gcatlataatgagagca.....ataaatgtcatgaccat 1126

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US10010580/runat_27062003.104535.11721/app-query.fasta.1.1287
-DB=SwissProt.40 -OEXT=fastan -SUFFIX=2p.rsp -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010580_@CGN_1_1_26_@runat_27062003.104535.11721 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEBUFFER -NEG_SCORES=0 -WAIT -DSPHLOCK=100 -LONCLOCK
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1568	78.5	331	CATS_HUMAN	P25774 homo sapien
2	1367.5	68.5	340	CATS_MOUSE	O70370 mus musculu
3	1334.5	66.8	330	CATS_RAT	Q03765 rattus norv
4	1089	54.5	329	CATS_BOVIN	P25326 bos taurus
5	944.5	47.3	329	CATK_RABIT	P43236 oryctolagus
6	943.5	47.2	329	CATK_MACRA	O77641 macaca fasc
7	939.5	47.0	329	CATK_HUMAN	P43235 homo sapien
8	937.5	46.9	329	CATK_RAT	O35186 rattus norv
9	933.5	46.7	330	CATK_PIG	O941e1 sus scrofa
10	926.5	46.4	329	CATK_MOUSE	P50597 mus musculu
11	843	42.2	334	CSL2_HUMAN	O60911 homo sapien
12	835.5	41.8	334	CATL_MOUSE	P06797 mus musculu
13	823.5	41.2	334	CATL_RAT	P07154 rattus norv
14	812	40.7	334	CATL_BOVIN	P25975 bos taurus
15	806	40.4	333	CATL_HUMAN	P07711 homo sapien
16	804	40.3	334	CATL_PIG	O28944 sus scrofa
17	777	38.9	323	CYS2_HUMAN	P25782 homarus ame
18	767	38.4	333	CATL_MOUSE	O94014 mus musculu

19	767	38.4	341	CATL_DROME	U95029 drosophila
20	759	38.0	321	CYS3_HOMAM	P25784 homarus ame
21	748	37.5	333	TEST_RAT	P15242 rattus norv
22	734	36.8	343	CATL_RAT	O947e3 rattus norv
23	715.5	35.8	322	CYS1_HOMAM	P13277 homarus ame
24	689	34.5	344	CYS5_DICDI	P54640 dictyostell
25	685.5	34.3	218	CATL_CHICK	P09648 gallus gall
26	675	33.8	458	ORYA_ORYSA	P25776 oryza sativ
27	667	33.4	462	RD21_ARATH	P43297 arabidopsis
28	652	32.6	217	CATL_SHEEP	U10991 ovis aries
29	647	32.4	471	ORYB_ORYSA	P25777 oryza sativ
30	634	31.7	395	CATL_BROPA	O17473 brugia paha
31	617.5	30.9	376	CYS2_DICDI	P04989 dictyostell
32	616.5	30.9	328	CYS4_BRANA	P25521 brassica na
33	603	30.2	380	ACTN_ACTCH	P25785 actinidia c
34	592.5	29.7	362	CYS1_PPAUV	P25803 phasecolus v
35	587	29.4	371	CYS1_HORVU	P25249 hordeum vul
36	586	29.3	373	CYS2_HORVU	P25250 hordeum vul
37	577	28.9	362	CYS1_VITGMU	P12412 vitina mungo
38	573.5	28.7	236	CATL_RAT	O63088 rattus norv
39	573	28.7	221	GPII_ZINOF	P82474 zingiber of
40	565	28.3	360	CYS2_MALIZE	O10717 zea mays (m
41	564	28.2	360	CYS1_HEMSP	P43156 hemerocalli
42	563	28.2	346	CYS1_LYCCE	P20721 lycopersico
43	561.5	28.1	333	CATL_RAT	P00786 rattus norv
44	559.5	28.0	345	ANAN_ANNACO	P80884 ananas como
45	558	27.9	221	GPI_ZINOF	P82473 zingiber of

ALIGNMENTS

RESULT 1

ID	CATS_HUMAN	STANDARD;	PRT;	331 AA.
AC	P25774; Q9BUC3;			
DT	01-MAY-1992 (Ref. 22, Created)			
DT	15-JUN-2002 (Ref. 41, Last sequence update)			
DT	15-JUN-2002 (Ref. 41, Last annotation update)			
DE	Cathepsin S precursor (EC 3.4.22.27).			
GN	CTSS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Alveolar macrophage;			
RX	MEDLINE=92218373; PubMed=1373132;			
RA	Shi G.-P., Munger J.S., Meara J.F., Rich D.H., Chapman H.A.;			
RT	"Molecular cloning and expression of human alveolar macrophage			
RT	cathepsin S, an elastolytic cysteine protease.";			
RL	J. Biol. Chem. 267:7258-7262(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94209337; PubMed=8157683;			
RA	Shi G.-P., Webb A.G., Foster K.E., Knoll J.H.M., Lemere C.A.,			
RT	Munger J.S., Chapman H.A.;			
RT	"Human cathepsin S: chromosomal localization, gene structure, and			
RL	tissue distribution.";			
RL	J. Biol. Chem. 269:11530-11536(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92317106; PubMed=1377692;			
RA	Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt H.,			
RT	Peters C.;			
RT	"Phylogenetic conservation of cysteine proteolases. Cloning and			
RL	expression of a cDNA coding for human cathepsin S.";			
RL	J. Biol. Chem. 267:13708-13713(1992).			
RN	[4]			
RP	REVISION TO 211.			
RC	TISSUE=Testis;			
RA	Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,			
RA	Peters C.;			

DB 321 GlycylalaserPhosphoSerTyProGluIle 331

RESULT 2

CATS_MOUSE STANDARD: PRT: 340 AA.
 ID CATS_MOUSE 070370; 054973;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calpessin S precursor (EC 3.4.22.27).
 GN CTSS OR CATS.
 OS Mus musculus (Mouse).
 OC Fukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 OX NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c, and 129/Sv; TISSUE-Brain;
 RA Doh-ura K.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Spleen;
 RA Romerskirch W.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 144-306 FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN-C57BL/6; TISSUE-Carilage;
 RA MEDLINE-99326135; PubMed-10395917;
 RA Soederstrom M., Salminen H., Glunoff V., Kirsche H., Aro H.,
 RA Vuorio E.;
 RT "cathepsin expression during skeletal development.";
 RL Biochim. Biophys. Acta 1446:35-46(1999).
 RN [4]
 RP SEQUENCE OF 296-340 FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-Brain;
 RA MEDLINE-98184882; PubMed-9516475;
 RA Dandoy-Dron F., Guillo F., Benoudjema L., Deslys J.-P., Laemesas C.,
 RA Dornont D., Tovey M.G., Dron M.;
 RT "Gene expression in scrapie. Cloning of a new scrapie-responsive gene
 and the identification of increased levels of seven other mRNA
 transcripts.";
 RL J. Biol. Chem. 273:7691-7697(1998).
 CC -1- FUNCTION: THIOL PROTEASE. THE BOND-SPECIFICITY OF THIS PROTEINASE
 IS IN PART SIMILAR TO THE SPECIFICITIES OF CATHEPSIN L AND
 CATHEPSIN N.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
 activity on Z-Phe-Arg-1-NHMeC, and more activity on the Z-Val-Val-
 Arg-1-Xaa compound.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION FOUND
 IN NON-SKELETAL TISSUES. RELATIVELY HIGH LEVELS FOUND IN SKELETAL
 TISSUES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF051732; AAC05781.1; -
 DR EMBL: AF051727; AAC05781.1; JOINED.
 DR EMBL: AF051728; AAC05781.1; JOINED.
 DR EMBL: AF051729; AAC05781.1; JOINED.
 DR EMBL: AF051726; AAC05781.1; JOINED.
 DR EMBL: AF051730; AAC05781.1; JOINED.
 DR EMBL: AF051731; AAC05781.1; JOINED.
 DR EMBL: AF038546; AAR84925.1; -
 DR EMBL: AJ002386; CAA05360.1; -

DR EMBL: Y18466; CAA77184.1; -
 DR EMBL: AJ223208; CAA11182.1; -
 DR HSSP: P25774; IEXF.
 DR MEROPS: C01.034; -
 DR MGD: MGI:107341; Ctss.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; Shprot_acstle.
 DR Pfam: PF00112; peptidase_C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR PRODOM: PD000158; Peptidase_C1; 1.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; lysosome; signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT PROPEP 18 112 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 113 340 CATHEPSIN S.
 FT ACT_SITE 147 147 BY SIMILARITY.
 FT ACT_SITE 287 287 BY SIMILARITY.
 FT ACT_SITE 307 307 BY SIMILARITY.
 FT DISULFID 134 233 BY SIMILARITY.
 FT DISULFID 144 189 BY SIMILARITY.
 FT DISULFID 178 222 BY SIMILARITY.
 FT DISULFID 281 329 BY SIMILARITY.
 FT DISULFID 120 120 N-LINKED (GLCNAC, . . .) (POTENTIAL).
 FT CARBOHYD 120 120 T->M.
 FT VARIANT 218 218 MRAPGHAIKRLFWMPVLCVYAMEQLORD -> MAVLDAPG
 FT VARIANT 1 29 VLGGNGATPER (IN REF. 2).
 FT CONFLICT 34 34 Y->H (IN REF. 2).
 FT CONFLICT 97 97 S->L (IN REF. 2).
 FT CONFLICT 106 106 S->P (IN REF. 2).
 FT CONFLICT 146 146 A->S (IN REF. 3).
 SU SEQUENCE 340 AA; 38438 MW; 06BE61126E2B0CDE CMC64;
 Alignment Scores:
 Pred. No.: 2,3e-114 Length: 340
 Score: 1367.50 Matches: 251
 Percent Similarity: 85.59% Conservative: 40
 Best Local Similarity: 73.82% Mismatches: 48
 Query Match: 68.48% Indels: 2
 Gaps: 1
 DB: 1
 US-10-010-580-1 (1-1126) x CATS_MOUSE (1-340)
 QY 12 ATGGGAGCAGCCGACAGTCTTCATGAATGGTAGTGGAGCTGCTCCGTTGCTGCTC 71
 DB 1 MetArgAlaProGlyHisAlaIleAlaGtrPleuPheTrpMetuProLeuValCysSer 20
 QY 72 TATGCAAGTGGCAGCAAGTACATGAAGCTCCGACCTGGATCATGCAATCTCTGCAG 131
 DB 21 ValAlaMetGluGlnLeuGlnArgAspProThrLeuAspTrpHisTrpAspLeuTrpLys 40
 QY 132 AAAAGCTACGCAAAACATC-AAGGAAGGAATGAGGAGTGCACAGGCTCAATG 190
 DB 41 LysThrHisGluLysGluTrpLysAspLysAsnGluGlnGlnValArgArgGluLeuTrp 60
 QY 191 GAAAAAATCTAAATTTGATGCTTCACATCTGGAACACTCAATGGGAATGATTC 250
 DB 61 GlutAsnMetuLysPheIleMetIleHisAsnLeuGluTrpSerMetGlyMetHisThr 80
 QY 251 TATGATCTAGGCAATGCAACCTCTGGAGACATGCTGGAAGTGAATGATCTTTCAG 310
 DB 81 TyrlValAlaIleMetAsnAspMetGlyAspMetLysAsnGluGlnLysSerCysArgMet 100
 QY 311 GGTTCCTGAGAGTTCCTCCAGCAATGGCAGAGAAATGTAACTTAAATCTGTANT 370
 DB 101 GlyAlaIleuArgIleSerArgInsSerProLysThrValThrPheArgSerTrpAsn 120
 QY 371 CAGAAATTCCTGATTTCTGTGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
 DB 121 ArgThrLeuProAspThrValAspTrpArgGluLysGluCysValThrGluValLysTrp 140
 QY 431 CAGGTTCTGTGTGTGCTTGTTGGGGCTTTCAGCGCTCTGGGGCCCTCGAAGCAGCAAC 490

```

DB 141 GINGLYSERCYSGLIALACYSSTRIPALAPHESERIALVALGILALALEUGLIGLINDLEU 160
UY 441 AAGCTAAACAGACGAAAGCTGCTGCTCTGAGTCAGACAGAACTTGATGCTCAACT 550
DB 161 LYSLEULYSTHGLYLSEULELESELESESERIALAGINALEVALASPCYSERASN 180
UY 551 GAAAAATATGGCATTAAGGCTGCAATGACGCTTTCATGACAACTGGTTCACAT 607
DB 181 GUGUGUGUSTYRGLYASNUYSGLYCYSGLYGLYRMEITHNGUUALAPHEGLNTRY 200
UY 608 ATTATGATTAACAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667
DB 201 LLELEASPASUGLYGLYILEGLUALASPALASERTRYPROTYRGLYALATHASPLU 220
UY 668 AAGTCACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
DB 221 LYSCTSYHSTYFASBSERLYSASNAFGLALATHRCYSSERAAQYRILEGLINLEUPRO 240
UY 728 TTGGCAATTAACATATTTTAAAGAGAGTGTGCAATTAAGAGAGCTGTGTGTGTGT 787
DB 241 PHEGLYASPGIUALASPALALACULYSGLUALAVAIATATHRYSGIYPROVALSERVALC 260
UY 788 ATACATGAGAGAGATATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847
DB 261 LLEASPALASERHISSESESESESESESESESESESESESESESESESESESESESE 280
UY 848 TCTACATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907
DB 281 CYSYTHRGILYASNAVALASNHISGLYVALLEUVALVALGILYRGLYTHLEUASPGI 300
UY 908 GACTAGCTGCTTTGTAAGAAAGCTGGGGCTTCAACTTTGGTGAACGAGATATATTCGG 967
DB 301 ASPTYRTPLEUVALIYASASNSERTPRCPILYEUASNPHEGLYASPGIINCIYTYRI 320
UY 968 ATGCTACAGAAATATGAAATATGTTGGATGATGATGATGATGATGATGATGATGAT 1027
DB 321 METALALAGASASALYASASNHISCYSGILYILEALASERTYRCSSETRYPROGLIU 340

```

RESULT 1
 CATS_RAT
 ID CATS_RAT STANDARD: PRT: 330 AA.
 AC 002765;
 DT 01-JUL-1994 (Ref. 26, first sequence update)
 DT 01-JUL-1994 (Ref. 26, last sequence update)
 DT 15-JUN-2002 (Ref. 41, last annotation update)
 DE Cathepsin S precursor (EC 3.4.22.27).
 GN CTSS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI-TaxID=10116;
 RN 111
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94100427; PubMed=1281481;
 RA Petancska S., Devi L.;
 RT *Sequence analysis, tissue distribution, and expression of rat
 cathepsin S.*
 JT Cathepsin S.;
 RL J. Biol. Chem. 267:26048-26043(1992).
 CC -1- FUNCTION: THIOLE PROTEASE. THE BOND-SPECIFICITY OF THIS PROTEINASE
 IS IN PART SIMILAR TO THE SPECIFICITIES OF CATHEPSIN L AND
 CATHEPSIN N. MAY BE INVOLVED IN THYROID HORMONE BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
 activity on Z-Phe-Arg-1-NHMe, and more activity on the Z-Val-Val-
 Arg-1 Xaa compound.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS OCCUR IN THE LIVER FOLLOWED
 BY SPLEEN, BRAIN, THYROID, OVARY AND UTERUS. LOW LEVELS ARE
 FOUND IN THE LIVER, KIDNEY, JEJUNUM AND LUNG WITH LOWEST LEVELS
 IN THE HEART.
 CC -1- INDUCTION: BY THYROID-STIMULATING HORMONE

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DB EMBL: L03201; AAA40994.1;
DB PIR: A45087; A45087.
DB HSP: P25774; IBXF.
DB MEROPS: C01.034.-.
DB InterPro: IPR000668; Peptidase_C1.
DB InterPro: IPR000169; SHPTOL_acsite.
DB Pfam: PF00112; Peptidase_C1.
DB PRINTS: PR00705; PAPAIN.
DB ProDom: PD000158; Peptidase_C1.
DB PROSITE: PS00139; THIOLE PROTEASE_CYS_1.
DB PROSITE: PS00639; THIOLE PROTEASE_HIS_1.
DB PROSITE: PS00640; THIOLE PROTEASE_ASN_1.
DB K01: Hydroxylase; Thiol protease; Lysosome; Zymogen; Signal.
DB FT SIGNAL 1 17
DB FT PROPEP 18 112
DB FT CHAIN 113 330
DB FT ACT_SITE 137 137
DB FT ACT_SITE 277 277
DB FT ACT_SITE 297 297
DB FT DISULFID 124 222
DB FT DISULFID 134 179
DB FT DISULFID 168 211
DB FT DISULFID 271 319
DB FT CARBOHYD 100 100
DB FT CARBOHYD 110 110
DB SO SEQUENCE 330 AA, 36833 MW, 670E3F08D7749E9E CRC64;

```

Alignment Scores:
 Pred. No.: 2,03e-111 Length: 330
 Score: 1334.50 Matches: 257
 Percent Similarity: 85.54% Conservative: 27
 Best Local Similarity: 77.41% Mismatches: 44
 Query Match: 66.83% Indels: 5
 DB: 1 Gaps: 3

US-10-010-580-1 (1-1126) x CATS_RAT (1-330)

```

UY 41 ATGCTAGTAGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100
DB 1 METALALAVALLLEUGLIALAPROGLYVALLEUCYSASPSAGLYALATHRALGLUARG-PR 20
UY 101 CACTTGCATCATCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACT 159
DB 20 OTHLEUASPHISHTSTRAPLEUTRPLYSYSTRATGMEIARGASHTHTRAPCL 40
UY 160 GAATGAGGAGAGTACGACGCGGTCTCATCTGGGAAAAAATCTTAAATTTGTATGCTTGA 219
DB 40 NASNGUGUGIUALSPVALARGATGILEUTPRCPILYUASASNHISPHLEIEMCTLEUAI 60
UY 220 CAATCTGGAACACTCAATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 279
DB 60 SASNHLEUGLUNHISSESESESESESESESESESESESESESESESESESESESESESE 80
UY 280 CATGACTGCTGAAGAGATGATATCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 339
DB 80 PMETLHPROGLIUGLIVALLILEGLYRMEIYSELEUARGILEPROARGPROTRAPS 100
UY 340 GAGAAATGCTACTATATAGTCAACCTATATGAGAAATGCTGATGCTGATGCTGATGCTGAG 399
DB 100 NATGSEGLYTHLEULYSERSESESESESESESESESESESESESESESESESESESESE 120
UY 400 AGAGAGGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459

```



```

145 ValGlyAlaLeuGluGlyGlnLeuLysLysThrGlyLysLeuAsnLeuSerPro 164
UY CAGAACTGTGACATTCCTCACTCAAGAAATAATGCGAATAAAGCTGCAATGGCGTTTC 586
165 GlnAsnLeuValAspCysValSerGlu-----AsnAspGlyCysGlyGlyTyr 181
UY ATGCAACTGCTTTGCAATATATATTGATACCAAGCGCATGATGAGAAGCTTCCTAT 646
182 MetThrAsnAlaPheGlnTyrValGlnLysAsnArgGlyIleAspSerGluAspAlaTyr 201
UY CCGTCAACAAGCCATGCAATGCAAGACGATATGATCACTCAAAAAAGCCAGCTGCCATGT 706
202 ProTyrValGlnGlnGlnLysSerCysMetTyrAsnProThrGlyLysAlaAlaLysCys 221
UY TCAAGATATAGCAACTTCCTTGGCAGTCAGCATGCTTAAAGAAGCTGTGCCAAT 766
222 AtGlyTyrTATGAGTCTTCTGATATATATATATATATATATATATATATATATAT 241
UY AAGAGCTCTGCTTCTGATATATATATATATATATATATATATATATATATATATAT 826
242 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
UY GATGCTGATATGATCAATGCTCTGCTTCTGATATATATATATATATATATATATATAT 883
262 GlyValTyrTyrAspGlnSerCysAsnSerAspAsnLeuAsnAlaValLeuAlaVal 281
UY GCTATGCTCACTTATATGCAAGACACTGCTGCTTGAAGAAACAGCTGGCGCTCAAC 943
282 GlyTyrGlyIleGlnLysGlyAsnLysHisTyrIleIleLysAsnSerTyrGlyLys 301
UY TTGCTGTCAGCAAGGATATATGCGATGCGAAGAAAGTGGAAATCATCCTGGGATGCT 1003
302 TrrGlyAsnLysGlyTyrIleLeuMetAlaArgAsnLysAsnAlaCysGlyIleAla 321
UY 1004 AATTATGCTCTTCAAGCAAGATC 1027
DB 322 AsnLeuAlaSerPheProLysMet 329
RESULT 7
CATK_HUMAN STANJAKO; PRT; 329 AA.
AC P43235;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin K precursor (EC 3.4.22.38) (cathepsin O) (cathepsin X)
GN (Cathepsin O2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95104457; PubMed=7805878;
RA Shi G.-P., Chapman H.A., Bhatnari S.M., Delcuvu C., Reddy V.V.,
RA Weiss S.J.;
RT "Molecular cloning of human cathepsin O, a novel endoproteinase and
RT homolog of rabbit OC2.";
RL FEBS Lett. 457:129-134(1995).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=95118380; PubMed=7818555;
RA Tsuda T., Bilbe G., Ishibashi O., Tezuka K.-I., Kamegawa M.,
RA Koduka T.;
RT "Molecular cloning of human cDNA for cathepsin K: novel cysteine
RT proteinase predominantly expressed in bone.";
RL Biochem. Biophys. Res. Commun. 206:89-96(1995).
RN 13
RP SEQUENCE FROM N.A.
RC TISSUE=osteoclastoma;

```

```

RX MEDLINE=96172222; PubMed=8585423;
RA Li Y., Alexander M., Mucherplemign A.L., Yellick P., Chen W.,
RA Strashenko P.;
RT "Cloning and complete coding sequence of a novel human cathepsin
RT expressed in giant cells of osteoclastomas.";
RL J. Bone Miner. Res. 10:1197-1202(1995).
RN 14
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=96082523; PubMed=7576232;
RA Broemme D., Okamoto K.;
RT "Human cathepsin O2, a novel cysteine protease highly expressed in
RT osteoclastomas and ovary molecular cloning, sequencing and tissue
RT distribution.";
RL Biol. Chem. Hoppe-Seyler 376:379-384(1995).
RN 15
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=97185906; PubMed=9033587;
RA McGrath M.E., Klaus J.L., Barnes M.G., Broemme D.;
RT "Crystal structure of human cathepsin K complexed with a potent
RT inhibitor.";
RL Nat. Struct. Biol. 4:105-109(1997).
RN 16
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=98070737; PubMed=9405598;
RA Thompson S.K., Halbert S.M., Bossard M.J., Tomaszek T.A., Levy M.A.,
RA Zhao B., Smith W.W., Abdel-Meguid S.S., Janson C.A., D'Alessio K.J.,
RA McQueney M.S., Amegadzie B.Y., Hanning C.K., Jesberger R.L.,
RA Briand J., Sarkar S.K., Huddleston M.J., James C.F., Carr S.A.,
RA Ganes K.T., Shu A., Hays J.R., Bradbeer J., Zembryki D., Veker D.F.;
RT "Design of potent and selective human cathepsin K inhibitors that
RT span the active site.";
RL Proc Natl. Acad. Sci. U S A 94:14249-14254(1997).
RN 17
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF ZYMOGEN FORM.
RX MEDLINE=99110586; PubMed=9893980;
RA Lalonde J.M., Zhao B., Janson C.A., D'Alessio K.J., McQueney M.S.,
RA Orsini M.J., Debouck C.M., Smith W.W.;
RT "The crystal structure of human procathepsin K.";
RL Biochemistry 38:862-869(1999).
RN 18
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
RX MEDLINE=99156066; PubMed=10048321.
RA Sivaraman J., Lalumiere M., Menard R., Cygler M.;
RT "Crystal structure of wild-type human procathepsin K.";
RL Protein Sci. 8:283-290(1999).
RN 19
RP VARIANT PYNODYSOSTOSIS ARG-146.
RX MEDLINE=96355650; PubMed=8703060.
RA Gelb B.D., Shi G.-P., Chapman H.A., Desnick R.J.;
RT "Pynodysostosis, a lysosomal disease caused by cathepsin K
RT deficiency.";
RL Science 273:1236-1238(1996).
RN 20
RP VARIANT PYNODYSOSTOSIS VAL-277.
RX MEDLINE=98198344; PubMed=9529353;
RA Gelb B.D., Willner J.P., Dunn T.M., Kardon N.H., Verloes A.,
RA Benzin J., Desnick R.J.;
RT "Paternal uniparental disomy for chromosome 1 revealed by molecular
RT analysis of a patient with pynodysostosis.";
RL Am. J. Hum. Genet. 62:848-854(1998).
RN 21
RP FUNCTION: CLOSELY INVOLVED IN THE DISORDER OF BONE REMODELING.
CC PARTICIPATE PARTIALLY IN THE DISORDER OF BONE REMODELING.
CC POTENT ENDOPEPTIDASE ACTIVITY AGAINST FIBRINOGEN AT ACID PH. MAY
CC PLAY AN IMPORTANT ROLE IN EXTRACELLULAR MATRIX DEGRADATION.
CC CATALYTIC ACTIVITY: Broad proteolytic activity. With small-
CC molecule substrates and inhibitors, the major determinant of
CC specificity is p2, which is preferably Leu, Met > Phe, and not
CC Arg.
CC -1- SUBCELLULAR LOCATION: lysosomal.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OSTEOCLASTS (BONNS).
CC -1- DISEASE DEFECTS IN CTSK ARE THE CAUSE OF PYNODYSOSTOSIS, AN
CC AUTOSOMAL RECESSIVE OSTEOCLASTOPATHIA CHARACTERIZED BY

```



```

0X NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Histar; TISSUE=Bone;
RA Payne J.A., Miller L.R.;
CC - FUNCTION: Closely involved in osteoclastic bone resorption and may
CC participate partially in the disorder of bone remodeling. Displays
CC potent endopeptidase activity against fibrinogen at acid pH. May
CC play an important role in extracellular matrix degradation (by
CC similarity).
CC - CATALYTIC ACTIVITY: Broad proteolytic activity. With small-
CC molecule substrates and inhibitors, the major determinant of
CC specificity is p2, which is preferably Leu, Met > Phe, and not
CC Arg.
CC
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DK EMBL: AF010306; AAB65743.1;
DK HSSP: P43235; 7PCK.
DK MEROPS: C01.046;
DK InterPro: IPR000668; Peptidase_C1.
DK InterPro: IPR000169; SHprot_Lactid.
DK Pfam: PF00112; Peptidase_C1; 1.
DK PRINTS: PR00705; PAIRIN.
DK PRODOM: P000158; Peptidase_C1; 1.
DK PROSITE: P500139; THIOL_PROTEASE_CYS; 1.
DK PROSITE: P500639; THIOL_PROTEASE_HIS; 1.
DK PROSITE: P500640; THIOL_PROTEASE_ASN; 1.
DK Hydrolase: Thiol protease; zymogen; signal.
DK SIGNAL: 1
DK PROPEP: 16 114 ACTIVATION PEPTIDE.
DK CHAIN: 115 329 CATHESPIN K.
DK ACT_SITE: 139 139 BY SIMILARITY.
DK ACT_SITE: 276 276 HY SIMILARITY.
DK ACT_SITE: 296 296 HY SIMILARITY.
DK DISULFID: 136 127 BY SIMILARITY.
DK DISULFID: 170 210 BY SIMILARITY.
DK DISULFID: 269 318 BY SIMILARITY.
DK CARBOHYD: 103 103 N-LINKED (GLCNAC ) (POTENTIAL).
DK CARBOHYD: 213 213 N-LINKED (GLCNAC ) (POTENTIAL).
DK SEQUENCE: 429 AA; 36790 MW; H416008509CF606A CRC64;

Alignment Scores:
Pred. No.: 6,04e-76 Length: 329
Score: 937.50 Matches: 179
Percent Similarity: 70.73% Conservative: 53
Host Local Similarity: 54.57% Mismatches: 87
Query Match: 46.95% Indels: 10
DB: 1 Gaps: 4

US-10-010-580-1 (1-1126) x CATK_RAT (1-329)
UY 54 CTCTCTCGTTCCTCTCTATGCAAGTGCACAGCACTACATTAAGATCCCACTCTGATCAT 113
DB H LeuLeuProValIvalSerPheAlaLeuSer -----ProGluGluThrLeuAspThr 24
UY 114 CACTGGAAATCTCTGAAGAAAGCTACAGAAACATTAAGGAAGACAAAT GAGGAAGTA 172
DB 25 ClntrpGluLeuIprLysIleThrHisGlySerIleTrpAsnSerLysValAspGluIle 44
UY 173 GCAAGGCTGCATGATGGGAAAAAATTAATTAATTTGATGCTTACCAATCTGGACAC 232
DB 45 SerAlaGluLeuIleTrpGluLysAsnLeuLysIleSerValHisAsnLeuGluAla 64
UY 233 TCAATGGCAATGATCATTAATGATCATTAACCATGCAACCATCTGGACAGCATCTGTGAA 292

```

```

DB 65 SerLeuGluAlaHisThrIleTrpGluLeuAlaIleAsnHisLeuGlyAspMetThrSerGlu 84
UY 293 GAAGTGATATCTTGATGGTCCCTGAGAGTCCG-----AGCCAAATGGCAGCAAAAT 346
DB 85 GluValValGlnLysMetThrGlyArgValIprProSerArgSerPheSerAsnAsp 104
UY 347 GTCACATATGAGTCAAACTGTATTCAGAAATTCGCCGATTCCTGTGACGACAGCAAG 406
DB 105 ThrLeuIleThrProGluIleTrpGluGlyValProAspSerIleAspIleTrpLys 124
UY 407 GGCTGTGCTGCTGCAAGTGAATACAGAGGTTCTTGCTGTGCTGTGGCTTTCAACGCT 466
DB 125 GlyValValThrProValLysAsnGlnGlyCysGlySerIlePheSer 144
UY 467 GTGGGGCCCTGGAGAGACAACTGAAAGTAAAGAGGAAAGTGGTGTCTGATGTA 526
DB 145 AlaGlyAlaLeuGluGlnGlyLeuLysLysIleThrGlyLysLeuAlaLeuSerPro 164
UY 527 CAGAACTTGCTAGATCTCTCACTCACTGAAATAATATGCGATGAAGCGCTGTC 586
DB 165 GluAsnLeuValAspCysValSerGluAsnTrp-----GlyGlyGlyGlyTrp 181
UY 587 ATGACAACTCTCTTCCAGTATATTATTGATTAACACCGCATTCATTCACAGCTTCAT 646
DB 182 MetThrThrAlaPheGlnTrpValGlnGlnAsnGlyIleAspSerGluAspAlaTrp 201
UY 647 CCTCAAAAGCCATGATGAAGTGCAGATATGATGCTCAAAAAGGAGCTGCTCATAT 706
DB 202 ProIleValGlnGlnAspGlnSerCysMetTrpAsnAlaThrAlaLysAlaLysCys 221
UY 707 TCAAGATATACGTAACCTCCCTTGGCAGTGAAGATCCCTTAAAGACCTGTGGCAAT 766
DB 222 ArgGlyTrpArgGluIleProValGlyAsnGlnLysAlaLeuLysArgAlaValAlaTrp 241
UY 767 AAGGACCTGCTGCTGGCTATGATAGAGGAGCCACTATTCCTCTCTGTCAGACAAGT 826
DB 242 ValGlyIleProValSerValSerIleAspAlaSerLeuThrSerPheIlePheTrpSerArg 261
UY 827 GGATGCTACTATGTAACCATCTCTACTCAAG---MATGTAATCATGATGATTAAGT 883
DB 262 GlyValIleTrpLysArgLysGlnAsnGlnCysAspArgAspAsnValAsnHisAlaVal 281
UY 884 GCGTATGCTACCTTATATGGAAGACACTACTGCTGTGAAAAACAGCTGGCGCCGCAAC 943
DB 282 GlyTrpGlyLysGlnLysGlnLysLysTrpIleLeuLysAsnSerTrpLysGluSer 301
UY 944 TTGCTGACCAAGCATATATTCGATGCGCAAGCAAGCATGGAATTCATCTGGATTCCT 1003
DB 302 TrpGlyAsnLysGlyTrpValLeuLeuAlaIleArgAsnLysAsnAlaCysGlyIleThr 321
UY 1004 AGTTATCCCTTATCCCAAAATC 1027
DB 322 AsnLeuAlaSerPheProLysMet 329

RESULT 9
CATK_PIG STANDARD; PRT; 330 AA.
AC DQGLS3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Catepsin K precursor (EC 3.4.22.38).
GN CTSK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RC TISSUE=Thyroid;
RX MEDLINE=20534869; PubMed=11082042;
RA Tepel C., Broemme D., Herzog V., Bräx K.;

```

Cathepsin K in thyroid epithelial cells: sequence, localization and possible function in extracellular proteolysis of thyroglobulin.
 RT J. Cell Sci. 113:4487-4498(2000).
 CC - FUNCTION: Closely involved in osteoclastic bone resorption and may participate partially in the disorder of bone remodeling. Displays potent endoprotease activity against fibrinogen at acid pH. May play an important role in extracellular matrix degradation (By similarity). Mediates thyroxine liberation by limited proteolysis of thyroglobulin at neutral pH in vitro.
 CC CATALYTIC ACTIVITY: Broad proteolytic activity. With small-molecule substrates and inhibitors, the major determinant of specificity is P2, which is preferably Leu, Met > Phe, and not Arg.
 CC SUBCELLULAR LOCATION: Lysosomal and extracellular.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF292030; MAG13340.1.
 DR DR HSSP: P43235; IMEM.
 DR DR MEROPS: C01.036; -.
 DR DR Interpro: IPR000668; Peptidase_C1.
 DR DR Interpro: IPR000169; SHPTC_acsite.
 DR DR Pfam: PF00112; Peptidase_C1; 1.
 DR DR PRINTS: PRO00705; PAPA1N.
 DR DR ProDom: PD000158; Peptidase_C1; 1.
 DR DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
 DR DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 DR DR HydroLase: Thiol protease; Lysosome; Glycoprotein.
 KM SIGNAL: 1 16 POTENTIAL.
 FT PROPEP 17 115 ACTIVATION PEPTIDE.
 FT CHAIN 116 330 CATHEPSIN K.
 FT ACT_SITE 140 140 BY SIMILARITY.
 FT ACT_SITE 277 277 BY SIMILARITY.
 FT ACT_SITE 297 297 BY SIMILARITY.
 FT DISULFID 137 178 BY SIMILARITY.
 FT DISULFID 171 211 BY SIMILARITY.
 FT DISULFID 270 319 BY SIMILARITY.
 FT CARBOHYD 104 104 N-LINKED (GLCNAC... (POTENTIAL).
 SU SEQUENCE 330 AA; 37069 MW; ABF089EF5FFB170 CRC64;
 Alignment Scores:
 Pred. No.: 1.38e-75 Length: 330
 Score: 933.50 Matches: 183
 Percent Similarity: 70.39% Conservative: 50
 Best local Similarity: 55.29% Mismatches: 89
 Query Match: 46.75% Indels: 10
 DB: 1 Gaps: 4
 US-10-010-580-1 (1-1126) x CATK_P1G (1-330)
 QY 45 CTAGTTGGCTGCTCCCTGCTGCTCATAGTGGCAGACAGTACATAAAGATCCCACT 104
 Db 6 ValValLeuLeuLeuProValMeIseSerFAla-----LeuTYrProGluGluIle 22
 QY 105 CTGATCATCATCGAATTCCTGGAAGAAAACCTACAGCAACAAATCAACAGACAGAT- 163
 Db 23 LeuAspThrGlnTrpGluLeuTrpGluLysLysThrTyArgLysGlnTyAsnSerLysVal 42
 QY 164 GAGAGATAGACGCGCTCATCTGGAAGAAAATCTAAATTGTGATGCTTCACAAAT 223
 Db 43 AspIuIleSerATArgLgLeuIleTrpGluLysLysLysLysLysLysLysLysLysLys 62
 QY 224 CTGGAACACCTGATGCAATGATATGATATGATGATGATGATGATGATGATGATGATG 283
 Db 63 LeuGluAlaSerLeuGlyValHisThrTyGluLeuAlaMetAsnHisLeuGlyAspMet 82

QY 284 ACTGTCAGACAGATGATATCTTGATGGGTCCCTGAGAGTTCCTCAAGCAAGATGAGAGA 343
 Db 83 ThrSerGluIuValValGlnLysMetThrGlyLeuLysValProProSerHisSerArg 102
 QY 344 -----AATGCTATTATAGTCCMACTCTATACAGAAATGCTGATTCTGTAATGAG 397
 Db 103 SerAsnAspThrLeuTrpLeuProAspTrpGluGlyArgThrProAspSerHisAspTy 122
 QY 398 AGACAGACAGCGCTGTCTTACAGCAAAATACACAGCGTCTCTGCTGCTGCTGCTGCTG 457
 Db 123 ArgLysLysGlyTyArgValThrProValLysAsnGlnGlyGlnGlySerGlySerPhe 142
 QY 458 TTCAGCGCTGTGGGGGCGCTGGAAGCAGACAGCTAAGCTAAACAGAGAAAGCTGTGCT 517
 Db 143 PheSerSerValGlyAlaLeuGluGluGlnLeuLysLysLysThrGlyLysLeuLeuAsn 162
 QY 518 CTGATGTCACAGAACTTGATGATGCTCACTGCAATGAAAATATGGAATTAAGCTGCAAT 577
 Db 163 LeuSerProGlnAsnLeuValAspCysValSerGlu-----AsnAspGlyCysGly 179
 QY 578 GCGGCTTCATGACACAGCTGCTTCAGTATATATGATTAACAACCGCATTCATTCAGAA 637
 Db 180 GlyLysTyMetThrAsnAlaPheGlnTrpValGlnLysAsnArgGlyLeuAspSerGlu 199
 QY 638 GCTTCCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGACTCAAAAAGACGACT 697
 Db 200 AspAlaTyTrpProTyArgValGlyGlnAspLysAsnGlyMetTyAsnProThrGlyLysAla 219
 QY 698 GCCCATGCTGTAAGTATGACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 757
 Db 220 AlaLysCysArgGlyTyArgGlnIleProGluLysLysLysLysLysLysLysLysLys 239
 QY 758 GTGGCCAAATGAAGACCTGTCTGTGGCTATATGATGCGACGCCACTATCTCTTCCTG 817
 Db 240 ValAlaArgValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPhe 259
 QY 818 TACAGATGCTGCTCTACTATGAAACCATCCCTG---ACTGAAATGTCAGATGAGATA 874
 Db 260 TyrSerLysGlyValTyTrpAspLysAsnGlySerAspAsnLeuAsnHisAlaVal 279
 QY 875 TTAGTGTGCTGCTATGCTTACCTTAATGGAAGAAAGCTACTGCTGCTGCTGCTGCTG 934
 Db 280 LeuAlaValGlyTyGlyIleGlnLysGlyLysLysHisTrpIleLeuLysAsnSerTrp 299
 QY 935 GCGCTCAACTTGTGCTGACCAAGATATATTCGATGCGCAACAAACAGTCAATACACT 994
 Db 300 GlyLysAsnTrpGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 319
 QY 995 GGGATTGCTAGTTATCCCTTACCCAGCAAAATC 1027
 Db 320 GlyIleAlaAsnLeuAlaSerPheProLysMet 330
 RESULT 10
 CATK_MOUSE STANDARD; PRT; 329 AA.
 ID P55097; O88718;
 AC 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin K precursor (EC 3.4.22.38).
 GN CTSK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Calvaria;
 RX MEDLINE=96409328; PubMed=8814310;
 RA Rantakokko J.A., Aro H.T., Savontaus M., Vuorio E.;
 RT "Mouse cathepsin K: cDNA cloning and predominant expression of the gene in osteoclasts, and in some hypertrophic chondrocytes during


```

SP
SEQUENCE OF 114-132 AND 293-311.
RC
TISSUE=Kidney:
RX
MEDLINE-92384944: Pubmed-1515067:
RA
Doelenc I., Rittmouja A., Collic A., Podobnik M., Oginc T., Turk V.;
" Bovine cathepsin S and L: Isolation and amino acid sequences.";
RL
Hiol. Chem. Hoppe-Seyler 373:407-412(1992).
CC
-1- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
LYSOSOMES.
CC
-1- CATALYTIC ACTIVITY: Specifically close to that of papain. As
compared to cathepsin B, cathepsin L exhibits higher activity
towards protein substrates, but has little activity on Z-Arg-Arg-
NHMeC, and no peptidyl-dipeptidase activity.
CC
-1- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
BONDS.
CC
-1- SUBCELLULAR LOCATION: Lysosomal.
CC
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
-----
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on
CC
use by non-profit institutions as long as its content is in no way
CC
modified and this statement is not removed. Usage by and for commercial
CC
entities requires a license agreement (see http://www.isb-slb.ch/announce/
or send an email to license@isb-slb.ch).
-----
DR
EMBL; X91755; CAA62870.1; -.
DR
EMBL; AB017648; BAA33398.1; -.
DR
EMBL; AB017645; BAA33398.1; JOINED.
DR
EMBL; AB017646; BAA33398.1; JOINED.
DR
EMBL; AB017647; BAA33398.1; JOINED.
DR
PIR; S15845; S15845.
DR
PIR; S24184; S24184.
DR
HSSP; 060911; 1FH0.
DR
MEROPS; C01.032; -.
DR
InterPro: IPR000668; Peptidase_C1.
DR
InterPro: IPR000169; SHprot_acsite.
DR
Pfam: PF00112; Peptidase_C1; 1.
DR
PRINTS; PR00705; PAPAIN.
DR
ProDom: PD000158; Peptidase_C1; 1
DR
PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR
PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR
PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KM
Hydrolase; Thiol protease; Lysosome; Zymogen; Signal.
FT
SIGNAL 1 17
FT
PROPEP 18 117
FT
CHAIN 118 289
FT
PROPEP 290 291
FT
CHAIN 292 334
FT
ACT_SITE 138 138
FT
ACT_SITE 277 277
FT
ACT_SITE 301 301
FT
DISULFID 135 178
FT
DISULFID 169 212
FT
DISULFID 270 323
FT
CARBOHYD 222 222
FT
CONFLICT 72 72
FT
CONFLICT 116 116
FT
CONFLICT 272 272
FT
CONFLICT 305 305
FT
CONFLICT 310 310
SQ
SEQUENCE 334 AA: 37377 MW: DABCF047DDEDEA2 CRC64:
-----
Alignment Scores:
Pred. No.: 9, 9e-65 Length: 334
Score: 812.00 Matches: 169
Percent Similarity: 61.67% Conservative: 45
Best Local Similarity: 48.70% Mismatches: 93
Query Match: 40.66% Indels: 40
DBs: 1 Gaps: 8
US-10-010-580-1 (1-1126) x CATL_BOVIN (1-334)

```

Db	7	LeuThrValLeuGlyValLeuGlnValAlaIleAspValAlaProLysLeuAspProAsnIleAsp	26
QY	111	CAIACAGTGAATCTCTGGAGAAACCTACAGCAAAATCAACAGAAAGCAATGAGGAG	170
Db	27	AlaHisThrPheIstGlnTrpLysAlaThrHisIleArgArg--LeuTYrGlyMetAsnIleGlu	46
QY	171	TAGCAGCGCCCTTCATCTGGCAAAAAATCTAAATTTTGATGCGTTCCACAATCTGCAC	230
Db	46	IuTrpArgGAlaValIleTrpGluLysAsnLysIleIleAspLeuHisAsnGlnGlu	66
QY	231	ACITCAATGGCAATGCATTCATATGATCTAGGCATGAAACCATCTGGGAGACATACCTG	290
Db	66	YrSerGluGlyLysHisAlaPheArgMetAlaMetAlaIlePheGlyAspMetHisAsnG	86
QY	291	AAGAAAGATATCTTTATGAGTGGCTCCCTGAGAGTCCAGTCCAAATGGCAGCAAAATGTCA	350
Db	86	IuGluPheArgGlnValMetAsnIlePheGln-----	96
QY	351	CTTATAGGTCAAACTGTAATCAGAA-----	376
Db	97	-----AsnGlnLysHisIstLysGlyLysLeuPheHisGluProLeuL	111
QY	377	-----TTGGCTGATCTGTGGCTGAGCAGCAAGAACGCCCTGTGATCACTGAAT	428
Db	111	euaValAspValAlaProLysSerValAspTrpTrpLysGlyTyrValIleThrProValLysA	131
QY	429	ACCAAGGTTCTTGTGTGCTGTGTGGCTTTCAAGTGTGTGGAGAGCTTGGCAACAC	488
Db	131	sngInGlyLncysGlySerCysTrpAlaPheSerAlaThrGlyAlaIleuGluGlyGln	151
QY	489	TGAACGTAAACACGAAAGCTGGTCTGCTGATGCGACACAACTGGTACATGGTCA	548
Db	151	eIlePheArgLysThrGlyLysLeuValSerLeuSerGluGlnAsnLeuValAspLysSerA	171
QY	549	CTGAATAATATGGGAATAAAGCTGCATGCGGCTTTCATGACCAACTGCTTTCCAGTATA	608
Db	171	rgAlaGln--GlyAsnGlnLysCysAsnGlyLysLeuMetAspAsnAlaPheGlnTyrI	190
QY	609	TTATTTGATTAACAAGGCAATGATATTCACAGAAGCTTCCTTCCCTACAAAGCATGAA	665
Db	190	IeLysAspAsnGlyGlyLysAspSerGluLysTyrProTyrLeuAlaThrAspThrA	210
QY	666	GAAATGTCAGATATGACTCAAAAAAGCGAGCTGCCACATGTTCAAAGTATACTGAATTC	725
Db	210	snsrCysAsnTyrLysProGluCysSerAlaIleAsnAspThrGlyPheValAspIleP	230
QY	726	CCTTTGGCAGTGAAGATCCCTTAAAGACGCTGTGGCCATTAAGACGCTGTCTGTGCG	785
Db	230	ro--GlnArgGluLysAlaLeuMetLysAlaValAlaIleThrValIleProIleSerValA	249
QY	786	CTATGATGGCGAGCACTATCTTCTTCCTGTGATACAGAAGTGGTGTACTAGTGAACAT	845
Db	249	IaIleAspAlaGlyHisIstHisSerPheGlnPheTyrLysSerGlyIleTyrTyrAspProA	269
QY	846	CCTGTACT--CAGAATGTGAATCATGGAGTATTAGTGGTTCGCTAGTCT-----	892
Db	269	sprCysSerCysLysAspLeuAspHisGlyValIleuValIleValIleGlyLysPheGluGlyT	289
QY	893	--AACCTTAATGGGAAGACACTAGTGGCTTGTGCAAAAGCTGGCGGCTCAACTTTGGC	950
Db	289	hrAspSerAsnAsnLysPheTrpIleValLysAsnSerTrpIleProGluTrpGluTyr	309
QY	951	ACCAAGATATATTTATGATGGTGAAGAACATGTGAATATCTGTGATTTATTTATTTAT	1010
Db	309	rpAsnGlyTyrValLysMetAlaLysAspGlnAsnAsnHisCysGlyIleAlaThrAlaA	329
QY	1011	CCTCTTACCAGAAATC 1027	
Db	329	IaSerTyrProThrVal 334	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:18:34 ; Search time 104 Seconds
(without alignments)
4461.713 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1997
Sequence: 1 gcatattatgtagggagca.....aataatgctcatgacat 1126

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Pgapop 6.0, Pgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_n2p.model -DEV-xlp
-Q/can2.1/USPTO_spool/US10010580/runat_27062003.104536.11734/app.query.fasta.1.1287
-DB-SPTREMBL.21 -QFMT-fastan -SUFFIX-n2p.rspt -MINMATCH=0.1 -DOFCL=0
-LOOEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-Blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OFFMT-pio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USP=0810010580.acsn.1.1.138.runat.27062003.104536.11734 -NPU=6 -ICPU=3
-NO_MMP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONMLAS
-DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP ivirus:*
16: SP bacteriap:*
17: SP archcap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1375.5	68.9	340	Q99M14	Q99m14 mus musculus

2	1016.5	50.9	331	13	Q90324	Q90324 cyprinus ca
3	865.5	43.3	338	13	Q90MC2	Q90MC2 orcohyachu
4	835.5	41.8	334	11	Q90DC0	Q90DC0 mus musculus
5	835.5	41.8	334	11	Q91U20	Q91U20 mus musculus
6	825	41.3	338	5	Q9Y0X2	Q9Y0X2 artemia san
7	814.5	40.8	333	6	Q9GKX8	Q9GKX8 cerropithec
8	812	40.7	333	6	Q9GKX8	Q9GKX8 cerropithec
9	793.5	39.7	334	13	Q90686	Q90686 gailus gail
10	793.5	39.7	334	13	Q90686	Q90686 gailus gail
11	792.5	39.7	334	13	Q90686	Q90686 gailus gail
12	781	39.1	333	5	Q27708	Q27708 nephrops no
13	775.5	38.8	334	11	Q91XK6	Q91XK6 mus musculus
14	772	38.7	333	11	Q9J181	Q9J181 mus musculus
15	770.5	38.6	338	5	Q46030	Q46030 caenorhabd
16	768.5	38.5	337	5	Q45734	Q45734 caenorhabd
17	768	38.5	332	5	Q9NHD5	Q9NHD5 boophilus m
18	768	38.5	337	5	Q9NHD5	Q9NHD5 boophilus m
19	767.5	38.4	328	5	Q27759	Q27759 penaeus van
20	767	38.4	324	13	Q90499	Q90499 myxine glut
21	764	38.3	326	5	Q27760	Q27760 penaeus van
22	763.5	38.2	334	11	Q9J1A9	Q9J1A9 mus musculus
23	760.5	38.1	324	13	Q80PP6	Q80PP6 engraullis j
24	760.5	38.1	339	5	Q26636	Q26636 sarcophaga
25	760.5	38.0	344	5	Q26425	Q26425 bombyx mori
26	759	37.8	324	13	Q80FP7	Q80FP7 entraullis j
27	754.5	37.8	325	5	Q46152	Q46152 penaeus van
28	754	37.8	324	13	Q912F4	Q912F4 mus musculus
29	752.5	37.7	334	11	Q9ET52	Q9ET52 mus musculus
30	747	37.4	343	11	Q912F4	Q912F4 mus musculus
31	747	37.4	374	5	Q18455	Q18455 heterodera
32	736.5	36.9	353	5	Q18456	Q18456 heterodera
33	730.5	36.6	333	11	Q912P3	Q912P3 mus musculus
34	730.5	36.6	333	11	Q912P3	Q912P3 mus musculus
35	728	36.5	333	5	Q95VQ2	Q95VQ2 frankliniell
36	727.5	36.4	331	5	Q46031	Q46031 sitophilus
37	726.5	36.4	333	11	Q9J196	Q9J196 mus musculus
38	724.5	36.3	331	5	Q46032	Q46032 sitophilus
39	724.5	36.3	331	11	Q912F2	Q912F2 mus musculus
40	720	35.9	323	5	Q95VQ4	Q95VQ4 frankliniell
41	716.5	35.9	339	11	Q9J1B4	Q9J1B4 mus musculus
42	709.5	35.5	311	5	Q8T529	Q8T529 fasciola he
43	709.5	35.5	322	5	Q967D5	Q967D5 fasciola he
44	709.5	35.5	326	5	Q24940	Q24940 fasciola he
45	707.5	35.4	311	5	Q9GRW5	Q9GRW5 fasciola he

ALIGNMENTS

RESULT 1
Q99M14 PREDIMINARY; PRT; 340 AA.
ID Q99M14
AC Q99M14
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cathepsin S.
GN CTSS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002125; AA02125.1; -
DR HSSP; P25774; 18XF.
DR MEROPS; C01.034; -
DR MGD; MGI:107341; Ctss.
DR InterPro; IPR000668; Peptidase-C1.
DR InterPro; IPR000169; SHprot_accsite.
DR Pfam; PF00112; Peptidase-C1; 1.
DR PRINTS; PR00705; PAPAIN.

Prodom: p0000158; Peptidase_C1: 1.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN: 1.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS: 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS: 1.
 KW Hydrolase; Thiol protease.
 SO SQUENCE: 340 AA; 38456 MW; 076502611F319D87 CRC64;

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 1375.50	340	252
Percent Similarity: 86.18%	Conservative: 41	
Best Local Similarity: 74.12%	Mismatches: 46	
Query Match: 68.88%	Indels: 2	
	Gaps: 1	

US 10 010-580-1 (1-1126) x 099M14 (1-340)

```

OY 12 ATGGCAGCAGCTGACGCTCTTCATGAAATGGCTAGTGGCGCTTCCTGCTGCTGC 71
DB 1 MetArgAlaIArgGlyHisAlaAlaIleArgTrpLeuMetProLeuValCysSer 20
OY 72 TATGACTGTAATATTAAGATATCTGATGATGATGATGATGATGATGATGATGATG 131
DB 21 ValAlaMetGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40
OY 132 AAAACCTACAGCAACATC-AACGAAGACATGAGAGTACGACGCGCTTCATCTGC 190
DB 41 LysThrHisGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 60
OY 191 GAAAAAATCTAAATTTTGATGCTTCACATGTCGACACGACATGGAATGCAATGCAATCA 250
DB 61 GluLysAsnLeuLysPheMetIleMetIleHisAsnLeuLysSerMetCysGlyMetHis 80
OY 251 TATGATCTAGCATGAAACATCTGCGAGACATGACTGATGAAATGATATCTTGTATG 310
DB 101 G1ValAlaLeuArg11LeuProArgLysSerProLysThrValThrPheArgSerTyrSerAsn 120
OY 311 GCTTCCTGACAGCTTCCGACCAATGCGACAGAAATGCTATAGCTTAACCTCAAT 370
DB 371 CAGAAATTCGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
OY 431 CAGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
DB 141 GlnGlySerCysGlyAlaCysTrpAlaPheSerAlaValGlyAlaLeuGlnLysGlnLeu 160
OY 491 AACCTAAACAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
DB 161 LysLeuLysThrGlyLysLeuLysLeuSerLeuSerAlaGlnLeuValAspCysSerAsn 180
OY 551 - GAAAAATATGCAATTAAGCGCTGCAATGCGCTTCATGCAACTGCTTCGCAATAT 607
DB 181 GlnGluLysLysGlyAlaLysGlyCysGlyCysGlyCysGlyCysGlyCysGlyCysGly 200
OY 608 ATTATGATTAACAAATGATGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 667
DB 201 LysLeuLysThrGlyLysLeuLysLeuSerLeuSerAlaGlnLeuValAspCysSerAsn 220
OY 668 AACCTCAATATGCAATTAAGCGCTGCAATGCGCTTCATGCAACTGCTTCGCAATAT 727
DB 221 LysLeuLysThrGlyLysLeuLysLeuSerLeuSerAlaGlnLeuValAspCysSerAsn 240
OY 728 TTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
DB 241 PheCysLysPheLysPheLysPheLysPheLysPheLysPheLysPheLysPheLysPhe 260
OY 788 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 847
DB 261 LysAspAlaSerHisSerSerPhePhePhePhePhePhePhePhePhePhePhePhePhe 280

```

```

OY 848 TGTACTGCAATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907
DB 281 CysThrGlyAlaValAlaHisGlyValLeuValAlaGlyTyrGlyThrLeuAspLys 300
OY 908 GACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
DB 301 AspTyrTrpLeuValLysAsnSerTrpCysLeuAsnPheGlyAspGlnCysLysLeuArg 320
OY 968 ATGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1027
DB 321 MetAlaArgAsnLysAsnHisCysGlyLysAlaSerAspCysSerTyrProLysLeu 340

```

RESULT 2

PRELIMINARY: PRT: 331 AA.

```

ID 090324
AC 090324:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Cysteine proteinase.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_TaxId=7962;
RN 11
RF SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA Tsai Y.-J.; Huang F.-L.;
RT "Molecular cloning of a cysteine protease from carp ovary."
RL Submitted (MAR-1994) to the EMBL/Genbank/DBD databases.
DR EMBL: L30111; AAA49207.1; -.
DR HSSP: P25774; 1BXF.
DR MEROPS: C01.034; -.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHPTL_acstte.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAIN.
DR Prodom: p0000158; Peptidase_C1: 1.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN: 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS: 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS: 1.
KW Hydrolase; Thiol protease.
SO SQUENCE: 331 AA; 36229 MW; 076DF34D32C1CF82 CRC64;

```

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 1,89e-86 <td>331 <td>190</td> </td>	331 <td>190</td>	190
Percent Similarity: 1016.50 <td>Conservative: 48</td> <td></td>	Conservative: 48	
Best Local Similarity: 74.84% <td>Mismatches: 76</td> <td></td>	Mismatches: 76	
Query Match: 59.75% <td>Indels: 5</td> <td></td>	Indels: 5	
	Gaps: 3	

US-10-010-580-1 (1-1126) x 090324 (1-331)

```

OY 66 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 125
DB 12 CysSerAlaAlaLeuAlaPhePheSerTrpAsn-----LeuAspGlnHisTrpGluLeu 29
OY 126 TCGAAGAAACCTACAGCAACA-AATCAAGCAAGCAATGACAGCAAGCAAGCAAGCAAGCA 184
DB 30 TrpLysLysThrHisAsnLysPheTyrSerLysAspGlnGluLysGlnValArgArgGlu 49
OY 185 ATCTGGGAAAAAATCTAAATTTGTGATGCTTCAATCTGAAACACTCAATGAGAAATG 244
DB 50 LeuTrpGluArgAsnLeuGlyLeuIleThrLeuHisAsnLeuGluAspLeuHisGlyLeu 69
OY 245 CATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304
DB 70 HisSerTyrAspLeuGlyMetAlaHisMetGlyAspMetThrThrGlnGluLysLeuGln 89
OY 305 TTGATGAGCTTCTCTGACAGCTTCCAGCCCAATGCGACAGCAAAATGTCACCT--TATAGTCA 361
DB 305 TTGATGAGCTTCTCTGACAGCTTCCAGCCCAATGCGACAGCAAAATGTCACCT--TATAGTCA 361

```

Db 90 ThrleualathrlleatrgvalProproglyPheylsArgGlnthralaGlnPhevalGly 109
 QY 362 AACTTAATACAAATTCCTGATTCGTGACTGGAGAGAGAGAGGCTGTGTACTGAA 421
 Db 110 SerSerGlyAlaAlaIleProAspSerLeuAspIrrParglIulysGlyTyryValSerSer 129
 QY 422 CTGAATACACGAGCTTCCTGCGTGTGTGGCTTCACGCTTCGGGCGCCCTGAA 481
 Db 130 VallysanglnGlyAlaGlySerCysTyrralPheSerSerValGlyAlaGln 149
 QY 482 GCACAACTGAAGCTAAACAGAAAGCTGTGTCTGTGAGTCACAGAACTTGTAAT 541
 Db 150 GlyGlnLeuMetIysThrThrcGlyLysLeuValAspLeuSerProGlnAsnLeuValAsp 169
 QY 542 TGCTCAACTGAAAAATATGGGAATTAAGGCTGCAATGGCTTCATGACAACTGCTTC 601
 Db 170 CysSerSer-----SerTyrrGlyAsnTyrrGlySerGlyGlyLeuMetSerAlaAlaPhe 188
 QY 602 CAGTATATTATTAATACACGAGCTTCATGACAGCTTCCTGCTACCAAAAGCCAT 661
 Db 189 GlnTyrrValIleAspAsnGlyGlyIleAspSerGlnSerSerTyrrProTyrrGlnGlyVal 208
 QY 662 AATGGAAGTGCAGATATGACTCAAAAAGGAGCTGCCACATGTTTAAAGTATCTGAA 721
 Db 209 GlnGlyGlnCysArgTyrrAsnProSerGlnLeuAlaAlaAsnGlyThrIlyTyrrTyrr 228
 QY 722 CTTCCTTTGGCAGTGAAGATCCCTTAAAGAGCTGTGGCTTAAAGAGCTGTGTCT 781
 Db 229 ValAspGlnGlyAspGlnGlnAlaLeuLysGlnAlaValAlaAsnIleGlyProIleSer 248
 QY 782 GTGGCTATAGATGACGAGCCACTATCTTCTCCGTACAGAGCTGTGTACTATGAA 841
 Db 249 ValAlaIleAspAlaThrIleHisProGlnPheIleLeuTyrrArgSerGlyValTyrrAsnAsp 268
 QY 842 CCATCTCTACTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
 Db 269 ProSerCysThrThrAsnIleAsnIleAsnIleValIleValIleValIleValIle 288
 QY 902 GGGAAGACTACTGCTGTGTGAAAAACAGCTGGGCTCACTGAGTGGAGCAAGATAT 961
 Db 289 GlyGlnAspPheThrPheValIlyAsnSerTyrrGlyThrGlyPheGlyAspGlyIlyTyrr 308
 QY 962 ATTGGATGCAAGAAACAGTGCAGCTGTGGATGCTAGTTATCCCTCT 1015
 Db 309 IleArgMetAlaIleArgAsnGlnAsnAsnMetCysGlyIleAlaIleHisThrProSer 326
 RESULT 3
 Q90WC2 PRELIMINARY: PRT: 338 AA.
 ID Q90WC2
 AC 01-DEC-2001 (TREMBlrel 19, Created)
 DT 01-DEC-2001 (TREMBlrel 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel 20, Last annotation update)
 DE: Procathopsin L.
 OS: Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC: Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC: Proclanchiopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8022;
 RN 111
 RP SOURCE: FROM N.A.
 RC TISSUE: OVARY;
 RA Pral F., Kwon J.Y., Tyler C.R.;
 RT *Molecular characterization of putative yolk processing enzymes and
 RT their expression during oogenesis and embryogenesis in rainbow trout
 RT Oncorhynchus mykiss.*;
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF358668; AAK69706.1;
 DR InterPro: IPR000169; SHProl_acsite.
 DR InterPro: IPR000169; SHProl_acsite.
 DR Pfam: PF00112; Peptidase_C1.1;
 DR Pfam: PF000158; Peptidase_C1.1;
 DR ProDom: PD000158; Peptidase_C1.1;
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN; UNKNOWN_1.

DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 SQ SEQUENCE 338 AA: 38384 MW: 888C834D95D6CE1 CRC64;
 Alignment Scores:
 Pred. No.: 2,53e-72 Length: 338
 Score: 865.50 Matches: 175
 Percent Similarity: 67.07% Conservative: 49
 Best Local Similarity: 52.40% Mismatches: 99
 Query Match: 43.34% Indels: 11
 Gaps: 6
 US-10-010-580-1 (1-1126) x Q90WC2 (1-338)
 QY 48 GTTGGGCTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCT 104
 Db 6 LeuAlaValIleValLeuValLeuValSerAlaValCysAlaAlaProAlaArgPheAspSerGln 25
 QY 105 CTGCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 164
 Db 26 LeuGlnAspHisThrPheIleuThrPheLysAsnThrPheIleSerLysHis--TyrrHisGlnSerG 45
 QY 165 AGGAAGTACGACGGGCTCATCTGCGAAAAAACTTAAATTTGTGATGCTTCACATC 224
 Db 45 IuGlnGlyTyrrPargArgMetValIrrGlnLysAsnLeuLysLysIleGlnIleHisAsnL 65
 QY 225 TGAACACATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 284
 Db 65 euGlnHisThrMetGlyLysHisSerTyrrArgLeuGlnMetAsnHisPheGlyAspMet 85
 QY 285 CTGGGTGACAGATGATATCTTGATGGGTCCCTGACAGCTCCAGCCAAATGCCAG--A 341
 Db 85 hAsnGlnGlnPheArgGlnThrMetAsnGlyTyrrLysGlnThrThrcGlnArgLysPheL 105
 QY 342 GAAATGTCATTAATAGTCAAACTCAATCAAACTCAATCAAACTCAATCAAACTCAATCAAACT 401
 Db 105 ysdIySerLeuPheMetGlnProAsnTyrrLeuGlnAlaProLysAlaValAspIrrParg 125
 QY 402 AGAAGGCTGTGTACTGAAAGTGAATACCAAGGCTTGTGGCTGTGGCTGTGGCTGTGCA 461
 Db 125 IuysGlyTyrrValIrrProValIlyAspGlnGlySerCysGlnSerCysTyrralPheSer 145
 QY 462 GCCCTGTGGGGCCCTGGAAGCAACCTGAAGCTAAACAGCAAAAGCTGTGTCTGCA 521
 Db 145 erThrThrGlnAlaMetGlnGlnGlnPheArgLysThrGlyLysLeuValSerLeuS 165
 QY 522 GTGCACAGAACTGTGTGATGCTCAACTGAAATATGCAATTAAGCTGCATGCTG 581
 Db 165 erGlnGlnAsnLeuValAspCysSerArgProGln--GlnAsnGlnGlyCysAsnGlyG 184
 QY 582 GTTCATGACAACTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCTGCTTCCT 641
 Db 184 IlyLeuMetLysPheGlnAlaPheGlnIlyIleGlnAspAsnAlaGlyLeuAspIrrGlnGlnS 204
 QY 642 CTAATCCCTCAAAAGCCATGAAT--GGAAGTGCAGATATGACATCAAAAGCCAGCTG 698
 Db 204 erTyrrProTyrrValGlyThrAspLysProCysHisTyrrLysProGlnPheSerAlaA 224
 QY 699 CCACATCTCAAAATAGTACTGAACCTTCCTGCGAGTGAAGATGCTTAAAGAAAGCTG 758
 Db 224 IAsnGlnIuThrGlyPheValAspIleProSerGlyGlnHisIleMetIlyAsnAlaV 244
 QY 759 TGGCCATTAAGAGCTGT 818
 Db 244 AlAlaAlaValAlaGlyProValSerValAlaIleAspAlaGlnHisGlnSerPheGlnPheT 264
 QY 819 ACAGAGT 875
 Db 264 yrrGlnSerGlyIleTyrrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 284
 QY 876 TAGTGTTGCTATAGT-----AACCTTAATGGGAAAGCTATGCTGTGTGTA 923
 Db 284 euValIlyGlyTyrrGlyPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 304


```

QY 576 ATGGAGGCTTCATGACAACTGCTTTCCAGATATATATGATAACAAAGCGCATTCATGAG 635
    |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 179 snnglyglyleumetasptryalaphocintyvalalalspsnnglyglyleuaspserc 199
QY 636 AAGCTTCCTATCCCTACCAAGCAATGATGAAAGGAGATGATGCTAAAAAGCGAG 695
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 199 lugsertyrprotylgualathrcglucluglucysylstyrsanproglutyrserv 219
QY 656 CTCCACATGCTTCAAACTAGCACTGCTTTCCGCACTGCAAGATGCTTAAAAAGAG 755
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 219 alalaaasaptryhlypnevalaspllepro---lysgnnglyulysalaleumetlyaa 238
QY 756 CTGTGCTCAATTAAGCACTGCTTCTGTGCTATAGATGCGAGCACTATCTTTCTTCC 815
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 238 lavalalathvalcglyproilalalaleasplacglyhiscglucluserpneup 258
QY 816 TGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 258 hetyrlysglyglylylethyrpnecluproaspyserserlgluaspmetasphehly 278
QY 873 TATTAAGCTGCTGCTATGCT-----AACCTTAATGGAAGAGAGAGAGAGAGAGAG 920
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 278 alaevalvalcglytyltyltyltyltyltyltyltyltyltyltyltyltyltylty 298
QY 921 TGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 298 allysasnserttrypdyglucluglupglymetcglyglylyltyltyltyltyltylty 318
QY 961 GTGCAATACACTGAGATTCAGTATGCTATCCCTTACCCAGAAATC 1027
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 318 rgargasnhiscysclglylealaseralalasertryprothval 333

RESULT 8
P79722 PRELIMINARY: PRT: 336 AA.
ID P79722:
AC 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DF 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Cathepsin L.
GN Hggl.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCHL_TaxID-7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Vogel A., Gerster T.;
RT "Expression of a zebrafish cathepsin L gene in anterior mesoderm
RT and hatching gland."
RL Dev. Genes Evol. 206:477-479(1997).
DR EMBL: Y08321; CAA69623.1; -.
DR HISSP: 060911; JFH0.
DR MEROPS: C01.032.
DR ZFIN: ZDB-GENE:980526-285; hggl.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHProt_acsite.
DR Pfam: PF00112; Peptidase_C1.
DR PRINTS: PR00705; PAPA1N.
DR ProDom: PD000158; Peptidase_C1.1.
DR PROSITE: PS00640; THIOI_PROTHASE ASN: 1.
DR PROSITE: PS00139; THIOI_PROTHASE_CYS: 1.
DR PROSITE: PS00639; THIOI_PROTHASE_HIS: 1.
KW Hydrolase; Thiol protease.
SQ SEQUENCE 336 AA; 38029 MW; 5E795879F85F6FC2 CRC64;

```

Alignment Scores: 2,55e-67 Length: 336
 Pred. No.: 812.00 Matches: 161
 Score: 65.938 Conservative: 48
 Percent Similarity: 50.79% Mismatches: 98
 Best Local Similarity: 50.79%

```

Query Match: 40.66% Indels: 10
DB: 13 Gaps: 5
us-10-010-580-1 (1-1126) x P79722 (1-336)
QY 96 GATCCCACTCTGATCATCTACGTAATCTCTGAGAGAAAAGCAAGCAATCAAGG 155
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 21 Asplleclnleuaspasphtstpranserttrylsserghlhiscllysser--tyrh 40
QY 156 AAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 215
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 40 lsgluasvalcgluvalcglyargarmetlletpuglunleuadqvsllleglug 60
QY 216 TTCACATCTGGAGCACTGAGATGAGATGATGATGATGATGATGATGATGATGATGAT 275
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 60 lnhlsasnphecluttyrsertryclysahisthrpneysmctglymetasnglnphcg 80
QY 276 GAGACATGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 80 lypsmetlthrasnglugluphearglnalamelasnglylrylthlasaspproahng 100
QY 333 AATGCAAGAGAAATGCTACTATAGTCAAACTCTAATCAGAAATGCGCATTCCTGCG 392
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 100 lnhtrserglnglyprouleuphemetcluproserpnephelalalaproglnglvala 120
QY 393 ACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 120 sptrparglnglvalcglyvalcglyvalcglyvalcglyvalcglyvalcglyvalcgly 140
QY 453 GGGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 140 lpsrpheserterhrclyalaleugluclyglnleuphearglysthrghlyltyltyl 160
QY 513 TGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 572
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 160 lsermelsercglucluglnleuvalaspysserargplogln--glyasnglnlyc 179
QY 573 GGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 632
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 179 ysasnglyglyleuuelaspclnalapheglnttyrvalcglyluclyluclyluclyluc 199
QY 633 CAGAGCTTCCTATCCCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 199 erglucgluclusertryprotyrleualalargaspaspleuprocysargtyrpsproarq 219
QY 690 AGCGAGCTGCCACATGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 749
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 219 heasnavalalalysilertnrglypnevalaspilleproserglyasngluleuhalalum 239
QY 750 AAGAGCTGTGGCCAAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 809
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 239 clasnhalvalalalalalalalalalalalalalalalalalalalalalalalalal 259
QY 810 TCTTCCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 866
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 259 euglnpnehyrglnserglylyltyrtyltyltyltyltyltyltyltyltyltyltylty 279
QY 867 ATGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 914
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 279 lsalalalalalalalalalalalalalalalalalalalalalalalalalalalalal 299
QY 915 GCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 299 rlllevallylsasnserttryseraspplustpdylyaspysltyltyltyltyltylty 319
QY 975 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 319 ysasplysasnsnhiscysglyvalalalalalalalalalalalalalalalalalal 334

```

RESULT 9
 ID 09GL24 PRELIMINARY: PRT: 333 AA.
 AC 09GL24:

DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprol_acsite.
 DR Pfam: PF00705; PAPAIN.
 DR PRINTS: PR00705; PAPAIN.
 DR PRODOM: PD000158; Peptidase_C1: 1.
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN: 1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS: 1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS: 1.
 DR Hydrolyase: Thiol protease; Zymogen; Signal; Lysosome; Glycoprotein.
 KW SIGNAL: 1
 FT PROPEP: 1
 FT CHAIN: 120
 FT ACT_SITE: 144
 FT ACT_SITE: 281
 FT ACT_SITE: 301
 FT DISULFID: 141
 FT DISULFID: 175
 FT DISULFID: 274
 FT CARBOHYD: 108
 FT SEQUENCE: 334 AA: 37218 MW: DB3E23FBC21BF3B9 CRC64;

Alignment Scores:
 Prod. No.: 1,37e-65 Length: 334
 Score: 793.50 Matches: 162
 Percent Similarity: 62.80% Conservative: 49
 Best Local Similarity: 48.21% Mismatches: 116
 Query Match: 39.73% Indels: 9
 Gaps: 4

US-10-010-580-1 (1-1126) x Q90686 (1-334)

QY ATGAAATGGCTAGTGGCTGCTCCGTTGCTCCTATGAGCGGCAAGTACATATA 95
 DB 4 LeuHisTrpLeuAlaLeuLeuValLeuLeuLeuPrometAlaAlaAlaInleuAlaPro 23
 QY 96 GATCCCACTGCTGATCATCTGGAATCTCTGGAAGAAACCTACAGCAACATCAAG 155
 DB 24 GluProGluLeuAspAlaGlnTrpAspLeuTrpLysArgThrIleGlnLysAlaValGln 43
 QY 156 AAGCAATGAGCAAGTACGACGGCGCT---CATCTGGCAAAAAAATCTAAATTTGTGA 212
 DB 44 -ArgGlnGlyGlyArgAsnValProGluValAspLeuGlyGlnGluProGluValHisArg 63
 QY 213 TGCTTCAACATCTGAGCACTCAATGGAATGATTCATATGATTCATGATGACATC 272
 DB 63 GcysProGlnArgGlyAla-ArgLeuGlyLysHisSerPheGlnLeuAlaMetAsnIlyL 83
 QY 273 TGGGAGACATGATGCTGGAAGTATCTTGTGATGGTCCCTGAGAGTCCAGCC 332
 DB 83 euGlyAspMetLhrSerGlnGluValAlaArgThrMetThrGlyLeuArgValProArgS 103
 QY 333 AATGCCAGAGA-----AATCTCATATAGCTCAACCTTAATCAGAAATTGCCATTT 386
 DB 103 eArArgProArgProAsnGlyThrLeuTrpValProAspTrpSerArgAlaProAla 123
 QY 387 CTGTGAGCTGAGAGAGAGAGGCTGTGTTACTGAAAGTAAATCCAGAGTTCTTGTG 446
 DB 123 laValAspTrpArgArgGlyGlyValThrProValLysAspGlnGlyGlnGlyS 143
 QY 447 CTGTGAGCTGAGAGAGAGGCTGTGTTACTGAAAGTAAATCCAGAGTTCTTGTG 506
 DB 143 eArArgProArgProAsnGlyThrLeuTrpValProAspTrpSerArgAlaProAla 163
 QY 507 ACCTGCTGTCTGATGACAGCAAGTCTGATGATGCTCACTCACTCACTCACTCACT 566
 DB 163 yLeuLeuSerLeuSerProGlnAsnLeuValLysCysValSer-----AsnAsn 180
 QY 567 AAGCTGCAATGCGGCTTCATGACAACTGCTTCATGATATTTATGATAAAGCA 626
 DB 180 snGlyCysGlyGlyLysMetThrAsnAlaPheGlnLysValArgLeuAsnArgGlyI 200
 QY 627 TTGATTGAGAGCTTCCTATACCTACAAAGCATGAGTGAAGAGTGCAGATATGACTCA 686

DB 200 leAspSerGluAspAlaTrpProGlyIleGlyGlnAspLysCysMetLysSerProT 220
 QY 687 AAAACGAGCTCCACATGTTCAAGATATCTCACTCCCTTTGGCAGTGAAGATCCCT 746
 DB 220 hrGlyLysAlaAlaLysCysArgGlyLysArgGlyIleProGluAspAsnGlyLysAla 240
 QY 747 TAAAGAGCTGCTGGCCATTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
 DB 240 euLysArgAlaValAlaAlaArgIleGlyProValSerValGlyLysAspLysLeuProS 260
 QY 807 CTTCCTTCCTGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
 DB 260 eArPheGlnPheTrpArgLysArgLysValTrpLysPheGlnGlyAspProGluAsnLle 280
 QY 864 AATGAGCTGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 923
 DB 280 snHisAlaValLeuAlaValAlaGlyGlyLysAlaGlnLysGlyThrLysHisTrpLle 300
 QY 924 AAAACGCTGGGCTCAACTGTTGCTGACCAAGATATATCGAGTGCAGAAAGAGT 983
 DB 300 yAsnSerTrpGlyThrGlnTrpGlyAsnLysGlyLysValLeuLeuAlaArgAsnMet 320

984 GAATCAGCTGGGATTCCTAGTTATCCCTTACCCAGAAATC 1027
 DB 320 yGlnTrpCysGlyLysAlaAlaSerPheProLysMet 334

RESULT 11

OBMSH3 PRELIMINARY; PRT; 316 AA.

AC 01-MAR-2002 (TREMBLE, 20, Created)
 DT 01-MAR-2002 (TREMBLE, 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLE, 21, Last annotation update)
 DE Cathespine L-like cysteine protease.
 OS Rhodnius prolixus (Triatomid bug).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera;
 OC Heteroptera; Panheteroptera; Cimicomorpha; Reduviidae; Triatominae;
 OC Rhodnius.
 NCBI_TaxID=13249;

RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=INTESTINE;
 RA Lopez-Ordóñez T., Rodríguez M.H.;
 RT "Characterization of a cDNA encoding a cathespine L-like protein of
 RT Rhodnius prolixus.";
 RL Insect Mol. Biol. 10:505-511(2001).
 DR EMBL: AF320565; AAL34984.1;
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprol_acsite.
 DR Pfam: PF00705; PAPAIN.
 DR PRINTS: PR00705; PAPAIN.
 DR PRODOM: PD000158; Peptidase_C1: 1.
 DR PROSITE: PS00640; THIOI_PROTEASE_ASN; UNKNOWN_1.
 DR PROSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW protease.
 SO SEQUENCE. 316 AA: 35602 MW: B979F0FD1CE94722 CRC64;

Alignment Scores:
 Prod. No.: 1.68e-65 Length: 316
 Score: 792.50 Matches: 160
 Percent Similarity: 64.33% Conservative: 42
 Best Local Similarity: 50.96% Mismatches: 107
 Query Match: 39.68% Indels: 6
 Gaps: 3

US-10-010-580-1 (1-1126) x Q8WSH3 (1-316)

QY 99 CCCACTCTGAT-----CATCACTGCAATCTCTGAGAGAAACCTACAGCAACAA 149
 DB 4 ProserPheAspLleAspProGlnGluTrpLeuAlaPheLysAlaMetHisGlyLysAsn 23


```

Db      118  ysaaprrparqslugluclyltyvalthprowalalrqsndlnclyscysglyserc 138
QY      450  GTTGGCGCTTTCACCGCTGTGGGCGCCCTGGACGACACTGACGACCTAAAAACAGCAAAAC 509
Db      138  ystfrrlallphelaaialaiallglglglglglglglglglglglglglglglglglglgl 158
QY      510  TGGCTTTCFATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 569
Db      158  eutfrprouservalcghasnlleucaspyserser---lysthrvalcghasnllysg 177
QY      570  GCTGCAATGCTGCTTTCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 629
Db      177  lycsaglserserlythralahlsaglnalaphelglutltyvalleuysasnllysglyleu 197
QY      630  ATTGACAGCTTCCTTCATCCCTTCATCAAAAGCCATGACGACGACGACGACGACGACGACG 689
Db      197  lualaglnalathrlyfrrtlygllygllygllygllygllygllygllygllygllygl 217
QY      690  AAGCACTGCTGCAATGCTTCAAAATGATGACGACGACGACGACGACGACGACGACGACG 749
Db      217  snalaserlalsnlllethrasprryvalasnlleuopro---proasnlleucutyfleur 236
QY      750  AAGCACTGCTGCAATGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 809
Db      236  rpyvalalavalaslerlleglyprovalseralalalalalalalalalalalalalal 256
QY      810  TCTTCCTGTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
Db      256  heafprrhethrlyasnlllygllyllyllyllyllyllyllyllyllyllyllyllyl 276
QY      867  ATGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 914
Db      276  lsalalalalalalalalalalalalalalalalalalalalalalalalalalalal 296
QY      915  GAGCTGTCACAAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 974
Db      296  rpllellyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 316
QY      975  GAAACAGCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027
Db      416  ysaasphllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 333
RESULT 14
QY1411  PHELMINARY: PRT: 333 AA.
AC      09J181:
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      FPGS68 (Cathepsin 2 precursor).
CN      CTS6 OR FPGS68 OR CTS2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
OX      NCBI-TaxID:10090;
RN      111
SEQUENCE FROM N.A.
KA      Hombertger M.C., Himmelhauer H., Ruchmann J., Zeitz C., Fundele R.H.;
KA      "CDNA subtraction cloning reveals novel genes whose temporal and
KA      spatial expression indicates association with trophoblast invasion.";
KA      Dev. Biol. 0:0-(2000).
KT      121
RP      SEQUENCE FROM N.A.
KC      STRAIN-C57H1/6; ISSUUE-PLACENTA;
KA      Denning J.M., Konadio M., Reiman S., Werber I., Schwinde A.,
KA      Dandekar T., Peters C.;
KA      "Characterization of a cluster of cathepsin L-like cysteine peptidases
KA      and related genes on mouse chromosome 13.";
KT      Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
RL      EMBL: AF250840; AAF81277.1;
DR      EMBL: AY014780; AAK00509.1;
DR      HSSP: P07711; ICOL.

```

```

DR      MEMOPS: C01.031; -.
DR      MGD: MGI:1860275; Cts8.
DR      InterPro: IPR000668; Peptidase_C1.
DR      InterPro: IPR000169; Sprot_acste.
DR      Pfam: PF00112; Peptidase_C1; 1.
DR      PRINTS: PR00705; PAPAin.
DR      ProDom: PD000158; Peptidase_C1; 1.
DR      PROSITE: PS00639; THIOL_PROTEASE_HIS_1.
KW      Signal.
FT      SIGNAL      1      20      POTENTIAL.
FT      CHAIN      114      333      CATHEPSIN 2.
SQ      SEQUENCE      333 AA; 37803 MW; 57648FD720A9614E CRC64.
Alignment Scores:
Pred. No.:      1,4e-63      Length:      333
Score:      772.00      Matches:      157
Percent Similarity:      62.61%      Conservative:      54
Best Local Similarity:      46.59%      Mismatches:      106
Query Match:      38.66%      Indels:      20
DB:      Caps:      7
US-10-010-580-1 (1-1126) x 09J181 (1-333)
QY      54  CTGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107
Db      6  leuvalalaleucysyleuglyvalalaglvalthrlnseraserproserleu 25
QY      108  GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 167
Db      26  asperglutrrpnlnglutrpllyllyllyllyllyllyllyllyllyllyllyllylly 45
QY      168  AAGTAGCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
Db      45  lyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 65
QY      228  AAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
Db      65  luyrlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 85
QY      288  GTGCAAGAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 347
Db      85  lyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 100
QY      348  TCACCTTAGCTCAACCTCAATCAAGAA-----TTCCTGATTCGTGACT 395
Db      100  hearqllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 120
QY      396  GGAGAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 455
Db      120  rparqllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 140
QY      456  CTTTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
Db      140  lapseralalalaglvalalaglgllyllyllyllyllyllyllyllyllyllyllylly 160
QY      516  CTGAGTGGACAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 575
Db      160  roleuserthrlnasnlleuvalaspcyserserargleugl---glyasnllyllylly 179
QY      576  ATGGCGGTTTCATGACCACTGCTTCCAGTAATATATATATATATATATATATATATATAT 635
Db      179  helysglyserthrphelualalaleuylslyllyllyllyllyllyllyllyllyllyl 199
QY      636  AAGCTTCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 695
Db      199  luserthrlyfrrtlyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 219
QY      696  CTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
Db      219  lalialaglylthrserpheserphevaliser---asnsergllyllyllyllyllylly 238
QY      756  CTGTGGCAATTAAGAGCTGTGTGTGTGCTATATAGAGGACGACTATTCCTTCTCC 815

```



```

Db      238  laValaIaThrlleGlyProIleSerValGlylleAspAlaArgHisLysSerPheArgL 258
QY      816  TGTACAGAAAGTGTCTCTACTATGACCAATCTGACTCAGAAAT---GTGAATCATGAG 872
Db      258  eutYrATGGLUGlyllyeTYrGLuprPolysCysSerSerAsnIlelleasnhlsery 278
QY      873  TATTAGTGTGGCTATGTG-----AACCTTAATGGCAAGACTAGTGGCTGG 920
Db      278  alleValaValGlyTYrGLyTYrGLyGLySerAspGlyAsnLysTYrTrpleuI 298
QY      921  TGAAGAAACAGCTGGGCGCCCTCAACTTGTGACCAAGCATATATTCGATGGCAAGAAACA 980
Db      298  lelyAsnSerHisleGlylUGlnTrpGlyMetAsnGlyTYrMetLysLeuAlaArgGlyA 318
QY      981  GTGCAATACACTGTGGCATTTGCTACTTATCCCTCTTACCCGAAATC 1027
Db      318  rGAsnAsnHisCysGlylleAlaSerTYrAlaValTYrProArgVal 333

```

RESULT 15

```

O46030      PRELIMINARY:      PRT:      338 AA.
AC      01-JUN-1998 (TREMBLE: 06, Created)
DT      01-JUN-1998 (TREMBLE: 06, last sequence update)
DI      01-MAR-2002 (TREMBLE: 20, last annotation update)
DE      Cysteine proteinase.
OS      Sitophilus zeamais (maize weevil).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC      Cucujiformia; Phyllophaga; Curculionidae; Dryophorinae; Sitophilus.
OX      MCB_TaxID=7047;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97279038; PubMed=9133615;
RA      Matsumoto I., Emori Y., Abe K., Arai S.;
RT      Characterization of a gene family encoding cysteine proteinases of
RT      Sitophilus zeamais (maize weevil), and analyses of the protein
RT      distribution in various tissues including alimentary tract and germ
RT      cells.
RL      J. Biochem. 121:464-476(1997).
DR      FMBI; D82884; BAA2442.1; -.
DR      HSSP; 060911; IFHO.
DR      MEROPS; C01.092; -.
DR      InterPro; IPR000668, Peptidase_C1.
DR      InterPro; IPR000169, Shprol_acsite.
DR      Pfam; PF00112; Peptidase_C1.1.
DR      PRINTS; PR00705; PAPA1N.
DR      PRODOM; PD000158; Peptidase_C1.1.
DR      PROSITE; PS00640, THIOL_PROTEASE_ASN; 1.
DR      PROSITE; PS00139, THIOL_PROTEASE_CYS; 1.
DR      PROSITE; PS00639, THIOL_PROTEASE_HIS; 1.
KW      Hydrolase; Thiol protease.
SQ      SPOUNCE 338 AA, 37570 MW, 104E1C34CF262977 CRCL4;

```

Alignment Scores:

```

pred. No.      1.95e-63      length:      338
Score:          770.50      Matches:      167
Percent Similarity: 63.258      Conservative: 43
Best Local Similarity: 50.308      Mismatches: 110
Query Match:      38.58%      Indels:      13
DB:              5          Gaps:      5

```

US-10-010-580-1 (1-1126) x O46030 (1-338)

```

QY      56  GCTTCGCTGTGTCTCTCTATGCACTGGCACAAGTACATAAGATCCCATCTTGATCATCA 115
Db      8  AlaAlaValaValleSerCysGlnAlaValSer-PheTYrAsp---LeuValGlnGlnG 26
QY      116  CTGCAATCTCTGACAGAAACCTACAGCAACATC-AAGGAAGAGATGACGACGATAGC 174
Db      26  nTTPSerSerPheLysMetClnHisSerLysAsnTYrAspSerGluThrGlnGlnArgPh 46
QY      175  ACGGCGTCTCATCTGGGAAAAAATCTAAATTGTGATGCTTCACAAATCTGGACACTC 234

```

```

Db      46  eAlGMeLysIlePheMeClnAsnAlaHisLysValAlaLysHisAsnLysLeuPheSe 66
QY      235  AATGCGAATGATCTATATGATGATAGCATGAACATCTTCGGAGACATGACTGGTGAACA 294
Db      66  rGlnGlyPheValLysPheLysLeuGlyLeuAsnLysTYrAlaAspMetLeuHisleG 86
QY      295  AGTATATCTTTGATG-----GGTTCCTTGAGAGTTCTTCAGTCA 333
Db      86  upheValSerThrLeuAsnGlyPheAsnLysThrLysAsnAsnIleLeuLysCysLeuSe 106
QY      334  ATGGCAGAGAAATGTCACTATATAGTCAAACTTAATGAGAAATTCATATTCATGGA 393
Db      106  pleuAsnAspAlaValaLArgPheIleSerProAlaAsnValLysLeuProAspPhrValAs 126
QY      394  CTGAGAGAGAAAGGCGTGTACTACTAGCAGAAATACCAAGCTTCCTTGCTGGCTTCGG 453
Db      126  pTTPArgAspLysGlyAlaValThrGlnValLysAspGlnGlyHisCysGlySerCysTr 146
QY      454  GCGTTGAGCGCTGTGGGCGCCCTGGAGAGCAACTGAAGCTTAAGAAACAGAAAGCTGT 513
Db      146  pSerPheSerAlaThrGlySerLeuGlnGlyGlnHisPheArgLysThrGlyLysLeuVa 166
QY      514  GTCTGTGAGTGACACAGAACTTGTGATTCCTCAACTGAAATAATATGGAAATAAGCTG 573
Db      166  lSerLeuSerGlnGlnAsnLeuValAspCysSer---GlyArgTYrGlyAsnAsnGlyC 185
QY      574  CAATGGGCGGTTTCACAACTGCTTCGATATATATGATATGATACAAAGCAATGATTC 633
Db      185  sAsnGlyGlyLeuMetLysPheAsnAlaPheArgTYrIleLysAspAsnGlyllylIleAsp 205
QY      634  AGAAGCTTCCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGACTCAAAAAAGCG 693
Db      205  rGlnLysSerTYrProTYrLeuAlaGlnAspGlnLysCysHisTYrLysAlaGlnAsnSe 225
QY      694  AGTGGCAACATGTTCAAAGTATACTGAATCCCTTGCGAGAGTGAAGATGCTTAAGAGA 753
Db      225  rGlyAlaThrAspLysGlyPheValAspIleGlnGlnAlaAsnGlnAspAspLeuLysAl 245
QY      754  AGCTGTGGCCAAATGAAGGACCTGTGTCTGTGGCTATAGATGCGAGCCACTATTCTTCT 813
Db      245  aAlaValaIaThrValaGlyProValSerIleAlaIleAspAlaSerHisGlyThrPheG 265
QY      814  CCTGTACAGAAAGTGTCTCTACTATGAACCATCTGT---ACTGAGAAATGGAATCATGG 870
Db      265  nLeuTYrSerAspGlyValTYrSerAspProGlnCysSerSerGlnGlnLysAspHisG 285
QY      871  AGTATTAGTGTGTGCTATAGTAACTT---ATGAGAAAGCTATGAGTTGTGAAAAA 927
Db      285  yAlaLeuValaValGlyTYrGlyThrSerAspAspGlyGlnAspTYrTrpleuValLysAs 305
QY      928  CAGCTGGGCGCTCAACTTTGTGTGACCAAGCATATATTCGATGGCAAGAAACAGTGGAAA 987
Db      305  nSerTrpLysProSerTrpGlyLeuAsnGlyTYrIleLysMetAlaIaArgAsnGlnAspAs 325
QY      988  TCAGTGTGGATGTGACTTATTCCTCTTACCA 1021
Db      325  nMeCysGlyAlaIaSerGlnAlaSerTYrPro 336

```

Search completed: June 27, 2003, 11:33:13
 Job time : 115 secs



; INFORMATION FOR SEQ ID NO: 6

DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
QY 755 GCTGTGGCCATTAAGACCCCTGCTGCTGCTATACATGGACCCCATTTCTTCTTC 814
DB 241 AlaValAlaAsnLysCylProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTTCACAGAACTGGTGTCTACTATGAACATCTCTACTCAGAAATGTGAATCATGAGTA 874
DB 261 LeuTyrArgSerGlyValTyrTyrGlyProSerCysThrGlnAsnValAsnHisCylVal 280
QY 875 TTACTGTGTGTGTATGTGTAACCTTAATGGCAAGACACTACGCTGTGCAAAACAGCTGG 934
DB 281 LeuValValGlyTyrGlyValAspLeuAsnGlyLysGlyTyrTrpLeuValLysAsnSerTrp 300
QY 935 GGTCCACACTTTCGACAGACAGATATATGGCATGGCATGCAAGACAGTCGAATCACTGT 994
DB 301 GlyHisAsnProGlyGlyGluGlyTyrTrpLeuValLysAsnHisCys 320
QY 995 GGATTCGTAGTATATCTCTTACCCAGAAATC 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 4

US-08-915-095A-B
; Sequence 8, Application US/08915095A
; Patent No. 6384793
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTROCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D4
; CURRENT FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-915-095A-B

Alignment Scores:
Pred. No.: 4,950-166 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Host Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: 4 Gaps: 0

US-10-010-580-1 (1-1126) x US-08-915-095A-B (1-331)

QY 46 ATGAATGGCTAGCTTGGGCTGCTGCTGCTGCTATGCGACGACGACGACGACATATAA 95
DB 1 MetLysArgLeuValCysValIleLeuValCysSerSerAlaValAlaGlnIleCysHisLys 20
QY 96 GATCCCATCTTGATCTCATCTGATCTGATCTGTAATAAATACAGATGAAATATC-AAG 154
DB 21 AsprTriothricLeuAspHisHisThrHisLeuTrpLysLysThrTrpGlyLysGlnTyrLys 40
QY 155 CAACAGATGACGACAGTACCAACGCGGCTCATCTGCGAAATAAAATCTAAATTTTGATG 214
DB 41 GluLysAsnGluGluAlaValAlaArgIleLeuIleTrpLysAsnLeuLysPheValMet 60
QY 215 CTTCACATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTG 274
DB 61 LeuHisAsnLeuGluHisSerMetGlyMetHisSerTrpAspLeuGluGlyMetAsnHisLeu 80
QY 275 GGAGACATGATGATGATGAAGATGATATCTTATGAGTAACTGAGAGATTCAGAGCA 334
DB 81 GlyAspMetThrSerGlyGluValMetSerLeuThrSerSerLeuArgValIleProSerGln 100
QY 335 TGACAGAGAAATGCTACTTATAGCTCAAACTTAATCAGAAATTCCTGATTCCTGAC 394
DB 101 TrpIleAlaArgAsnIleThrTyrLysSerAspProAsnArgIleLeuProAspSerValAsp 120

QY 395 TGACAGAGAAAGGCTGTGTATTAATGAATATACAGATTTTATGATGTTCTTGG 454
DB 121 TrpArgGluLysGlyValThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
QY 455 GCTTCAGGCTGTGTGAGGCTGTGTAAGTAACTGAATGAATGAATGAATGAATGAATGA 514
DB 141 AlaPheSerAlaValIleValLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTCTAGTCGACACAGACTTGTGATCTGCTCAACTGAAAAATATGGCAATTAAGCTTC 574
DB 161 ThrLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlyAsnLysCys 180
QY 575 AATGGGCTTCATGACAGCTGCTTCCAGTATATATGATATGACAGAGCTGATTC 634
DB 181 AsnGlyCylPheMetThrThrAlaPheGlnTyrIleIleAspAsnLysGlyLysAspSer 200
QY 635 GAAGCTTCCTATGCTTACAAAGCTATGAATGGAAGATGATATATATGAAAAATGGA 694
DB 201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGCCACAGTTCAGATATACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 754
DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
QY 755 GCTGTGGCCATTAAGACCCCTGCTGCTGCTATACATGGACCCCATTTCTTCTTC 814
DB 241 AlaValAlaAsnLysCylProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTTCACAGAACTGGTGTCTACTATGAACATCTCTACTCAGAAATGTGAATCATGAGTA 874
DB 261 LeuTyrArgSerGlyValTyrTyrGlyProSerCysThrGlnAsnValAsnHisCylVal 280
QY 875 TTACTGTGTGTGTATGTGTAACCTTAATGGCAAGACACTACGCTGTGCAAAACAGCTGG 934
DB 281 LeuValValGlyTyrGlyValAspLeuAsnGlyLysGlyTyrTrpLeuValLysAsnSerTrp 300
QY 935 GGTCCACACTTTCGACAGACAGATATATGGCATGGCATGCAAGACAGTCGAATCACTGT 994
DB 301 GlyHisAsnProGlyGlyGluGlyTyrTrpLeuValLysAsnHisCys 320

RESULT 5

US-08-798-096-B
; Sequence 8, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTROCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D2
; CURRENT APPLICATION NUMBER: US/08-798,096
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-096-B

Alignment Scores:
Pred. No.: 4,950-166 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Host Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: 4 Gaps: 0

US-10-010-580-1 (1-1126) x US-08-798-096-B (1-331)

```

OY 36 ATGAATGGCTAGTGGAGCTGCTCCGTTGCTGCTATGCAAGTGGCAAGTACATMAA 95
Db 1 MetlysargleuvalCysvalleuvalCyssetserAlaValAlaGlnleuHislys 20
OY 96 GATCCCACTGATGATCAGTACCTGGAATCTCTGCAAGAAAACCTACAGCAACAAATC-AAG 154
Db 21 AsprothleuAspHisHisIstRPhIsleuTriPylsYsthrYleYlYsgIntrYlys 40
OY 155 CAACAGAAATGAGCAAGTACCAAGCGCTCATCTGCGAAAAAAATCTAAATTTGTGATG 214
Db 41 GlulysasnGluGluAlaValArgArgleuIleTriPolulysasnleuYspheValMet 60
OY 215 CTTCACAACTTGGAACACTCAATGGAAATGATTCATATGATAGCATGGAACCATGTG 274
Db 61 leuHisasnleuGluHisSerMetClyMetHisSerYthrAspleuGlyMetasnHisIleu 80
OY 275 GCAGACATGACCTGCGAAGTATATCTTGAATGGGTTCCCTGACAGCTCCACCCAA 334
Db 81 GlyAspMetThrSerGluGluValMetSerleuThrSerSerleuArgValProSerGln 100
OY 335 TGGCAGAAATGTCACTTATAGTCAAACTCTAATGAAATGGCTGATTTCTGGAC 394
Db 101 TrpGlnArgAsnIleThrYtrYlysSerasnProasnArgIleleuProAspSerValAsp 120
OY 395 TGGAGAGAGAGGGCTGTACTGAGTGAATACAGGGTTCTTGTGCTGTGG 454
Db 121 TrpArgGluYsgIlyCysValThrGluValIlystrGlnIlySerCysGlyAlaCysTrp 140
OY 455 GCTTTACGCCCTGGGGCCCTCGAAGCAACAGTAAAAACAGCAACCTGCTG 514
Db 141 AlaPheSerAlaValGlyAlaLeuGluAlaGlnleuYsgIleuYsthrGlyLysleuVal 160
OY 515 TCTCTGAGTGCACAGAACTTGCTAGTCTCACTGAAATATATGCAATTAAGAGCTGCG 574
Db 161 ThrleuSerAlaGlnAsnleuValAspCysSerThrGluYsthrGlyAsnIlySclYlys 180
OY 575 AATGGCGTTTCATGAACTGCTTCCAGTATATGATATGATTAACAGCGCATTTGATCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnIlyIleleAspAsnIlySclYleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGAAATGCAAGTGCAGATATGACCTCAAAAACGA 694
Db 201 AspAlaSerYtrProYtrYlysAlaMetAspGlnIlySclYstrYleAspSerYstrYtrArg 220
OY 695 GCTGCCACATGTGAAATGATAGTCACTTCCCTTGGCAGTACAGATGCCCTTAAAGAA 754
Db 221 AlaAlaIthrCysSerYstrYtrThrGluLeuProIlytrGlyArgGluAspValleuYsgIu 240
OY 755 GCTGTGGCCATMAAGACTGTGTCTGTGCTATAGATGCGAGCGCATTTCTCTTC 814
Db 241 AlaValAlaAsnIlyGlyProValSerValGlyAlaAspAlaIthrGlnHisProSerPhePhe 260
OY 815 CTCTACAGAGTGGTGTACTATGAACCATCTCTACTGACAAATGGAATCATGAGTGA 874
Db 261 LeuIlyrArgSerGlyValIlytrGluProSerCysThrGlnAsnValAsnHisIlyVal 280
OY 875 TTAGTGGTGGCTATGCTTAACCTTAATGGAAAGCACTACTGCTTGCAAAAAACAGCTGG 934
Db 281 leuValIalGlyIlytrClyAspleuasnGlyIlystrYtrIlyPheValIlyAsnSerTrp 300
OY 935 GGGCTCACTTGTGACCAAGGATATATTGCGATGGCAAGAAACAGTGGAAATCACTGT 994
Db 301 GlyHisasnPheClyGlyGluGluIlytrIlyleArgMetAlaIthrAsnIlySclYasnHisCys 320
OY 995 GGCATTGCTAGTATCCCTTTACCCAGAAATC 1027
Db 321 GlyIleAlaSerPheProSerYtrProGluIle 331

```

RESULT 6
 US-08-798-095A-8
 : Sequence 8, Application US/08798095A
 : Patent No. 6423507
 : GENERAL INFORMATION:

```

: APPLICANT: Haslings, et al.
: TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
: FILE REFERENCE: PF107D3
: CURRENT APPLICATION NUMBER: US/08/798,095A
: CURRENT FILING DATE: 1997-02-12
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO: 8
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-08-798-095A-8

Alignment Scores:
Pred. No.: 4,95e-166 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Best Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
Gaps: 0

US-10-010-580-1 (1-1126) x US-08-798-095A-8 (1-331)

OY 36 ATGAATGGCTAGTGGAGCTGCTCCGTTGCTGCTATGCAAGTGGCAAGTACATMAA 95
Db 1 MetlysargleuvalCysvalleuvalCyssetserAlaValAlaGlnleuHislys 20
OY 96 GATCCCACTGATGATCAGTACCTGGAATCTCTGCAAGAAAACCTACAGCAACAAATC-AAG 154
Db 21 AsprothleuAspHisHisIstRPhIsleuTriPylsYsthrYleYlYsgIntrYlys 40
OY 155 CAACAGAAATGAGCAAGTACCAAGCGCTCATCTGCGAAAAAAATCTAAATTTGTGATG 214
Db 41 GlulysasnGluGluAlaValArgArgleuIleTriPolulysasnleuYspheValMet 60
OY 215 CTTCACAACTTGGAACACTCAATGGAAATGATTCATATGATAGCATGGAACCATGTG 274
Db 61 leuHisasnleuGluHisSerMetClyMetHisSerYthrAspleuGlyMetasnHisIleu 80
OY 275 GCAGACATGACCTGCGAAGTATATCTTGAATGGGTTCCCTGACAGCTCCACCCAA 334
Db 81 GlyAspMetThrSerGluGluValMetSerleuThrSerSerleuArgValProSerGln 100
OY 335 TGGCAGAAATGTCACTTATAGTCAAACTCTAATGAAATGGCTGATTTCTGGAC 394
Db 101 TrpGlnArgAsnIleThrYtrYlysSerasnProasnArgIleleuProAspSerValAsp 120
OY 395 TGGAGAGAGAGGGCTGTACTGAGTGAATACAGGGTTCTTGTGCTGTGG 454
Db 121 TrpArgGluYsgIlyCysValThrGluValIlystrGlnIlySerCysGlyAlaCysTrp 140
OY 455 GCTTTACGCCCTGGGGCCCTCGAAGCAACAGTAAAAACAGCAACCTGCTG 514
Db 141 AlaPheSerAlaValGlyAlaLeuGluAlaGlnleuYsgIleuYsthrGlyLysleuVal 160
OY 515 TCTCTGAGTGCACAGAACTTGCTAGTCTCACTGAAATATATGCAATTAAGAGCTGCG 574
Db 161 ThrleuSerAlaGlnAsnleuValAspCysSerThrGluYsthrGlyAsnIlySclYlys 180
OY 575 AATGGCGTTTCATGAACTGCTTCCAGTATATGATATGATTAACAGCGCATTTGATCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnIlyIleleAspAsnIlySclYleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGAAATGCAAGTGCAGATATGACCTCAAAAACGA 694
Db 201 AspAlaSerYtrProYtrYlysAlaMetAspGlnIlySclYstrYleAspSerYstrYtrArg 220
OY 695 GCTGCCACATGTGAAATGATAGTCACTTCCCTTGGCAGTACAGATGCCCTTAAAGAA 754
Db 221 AlaAlaIthrCysSerYstrYtrThrGluLeuProIlytrGlyArgGluAspValleuYsgIu 240
OY 755 GCTGTGGCCATMAAGACTGTGTCTGTGCTATAGATGCGAGCTTCTCTTC 814

```

Db	1	MethylarsenylvalylcysvalleucylcysserSerAlaValAlaGlnIleuHisLys	20
UY	96	GATCCACCTGTGATCATCTAGTGGATCTCTGGAGAAACCTACGTAACAACTG	154
Db	21	AspProIthrLeuAspHisHisIstIrrHisLeuIrrLysLysIrrLysLysLysLys	40
OY	155	GAAGACAACTAGCAACTAGCAGCGGCTGCATCTGGCAAGAAAAATCTAAATTTTCGATG	214
Db	41	GluLysAsnGluIleuAlaValAlaIrrGluLeuIleIrrGluLysAsnLysLysPheValMet	60
OY	215	CTTCACCAATCTGGAAATCTGAATGTATGATATATATATATATATATATATATATG	274
Db	61	LeuHisAsnLeuGlnHisSerMetGlyMetHisSerLysTrpAspLeuGlyMetAsnHisLeu	80
OY	275	GGAGCATATGCTGGTAAAGAACTGATATCTTTGATGGTTCCTCCGAGAGTCCCAAGCA	334
Db	81	GlyAspMetIrrSerIrrGluIleuValMetSerLeuMetSerSerLeuAlrrValProSerGln	100
OY	335	TGGCAGCAAAATGTCACTTATAGGTCAAACCTATATCAAGAAATTCGCGATCTGTCTGAC	394
Db	101	TrpGlnAlaGlnSerIrrLysSerIrrAsnProAsnAlaGlyLeuIrrProAspSerValAsp	120
OY	395	TGGAGAGAGAGAGGCTGTGTATTACTGAAAGTAAATATAGAGTTTATTGTGCTGTG	454
Db	121	TrpAlrrGluLysGlyValIrrGluValAlaLysTrpGlnLysSerCysGlyAlaLysTrp	140
OY	455	GCATTAGCCCTCTGGGGCCCTGGAGATATATATGAATTAATAATATATATATATAT	514
Db	141	AlaPheSerAlaValAlaIrrAlaLeuGluAlaGlnLeuLysLeuLysIrrGlyLysLeuVal	160
OY	515	TCTCTGACTGCACAGAACTTGGTACATTCCTCACTGAAAAATATGGCAATAAAGCTTCC	574
Db	161	SerLeuSerAlaGlnAsnLeuValAspCysSerIrrGluLysGlyAsnLysGlyLys	180
OY	575	AATGGGGTTCTTGCAACTGCTTTCCAGTATATATATATATATATATATATATATATATCA	634
Db	181	AsnGlyGlyPheMetIrrThrAlaPheGlnTrpIleIleAspAsnLysGlyLysLeuAspSer	200
OY	635	GAAGCTTCTATCTCTACAAAGCTCATGAATGGAAAGTGTATATGACTAAAAACGCA	694
Db	201	AspAlaSerLysProTrpLysAlaMetAspGlnLysGlnTrpLysSerLysLysLysLys	220
OY	695	GCATCCACATGTTCAAGATATACGTCAACTCTCCCTTGGCAGCTGAAGATGCTTTAAAA	754
Db	221	AlaAlaIrrCysSerLysTrpIrrGluLeuProTrpIrrAlaGrgGluAlaAspLeuLysGlu	240
OY	755	GCTGTGGCAATTAAGAGACTGTGTCTGTGGTATATATATATATATATATATATATAT	814
Db	241	AlaValAlaAsnLysGlyProValSerValGlyValAspAlaAlaGlnHisProSerPhePhe	260
OY	815	CTGTACAGAACTGGTGTAT	874
Db	261	LeuTrpAlrrSerSerLysValLysTrpGluProSerCysThrGlnAsnValAsnHisGlyVal	280
OY	875	TTTGTGCTGGCATATGCTAGCTTATGGGAAAGCACTGCGCTTGGAAGAAACAGCTTCG	934
Db	281	LeuValAlaLysLysIrrLysPheAsnGlnLysGlnTrpIrrPheValLysAsnSerTrp	300
OY	935	GGCGTCAACTTGGTGAATAGATATATATATATATATATATATATATATATATATATAT	994
Db	301	GlyHisAsnPheGlyGlnGlnGlyTrpIleArgMetAlaIrrAsnLysGlyAsnHisCys	320
OY	995	GGGATTCGTAATTATCTCTTACCCAGCAATC	1027
Db	321	GlyIleAlaSerPheProSerLysProGlnIle	331
RESULT 8			
PCT-US95-13820-4			
Sequence 4, Application PC/TUS9513820			
GENERAL INFORMATION:			
APPLICANT: Khepri Pharmaceuticals, Inc.			
TITLE OF INVENTION: CATHEPSIN O2 PROTEINASE			
NUMBER OF SEQUENCES: 12			

NAME: FERRARO, GREGORY D.
 REGISTRATION NUMBER: 46,134
 REFERENCE/DOCKET NUMBER: 425800-95
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 329 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS:
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 US-08-208-007A-7

Alignment Scores:
 Pred. No.: 5,260-97
 Score: 944.50
 Percent Similarity: 69.55%
 Best Local Similarity: 55.52%
 Query Match: 47.10%
 Matches: 186
 Conservative: 47
 Mismatches: 79
 Indels: 24
 Gaps: 5

US-10-010-580-1 (1-1126) x US-08-208-007A-7 (1-329)

```

UY 54 CTGATGGGTTGTTGTTATGTAATGCTAAATGATATGATGATCAT 113
DB 8 LeuLeuProValValSerPheAla-----LeuHisProGluGluLeuAspThr 24
UY 114 CACTGGCAATCTGTGTAAGAAAGAACTATGCAAGCAATGCAAGAGAGAT 172
DB 25 GlnTrpGluLeuIlePrlPylsLysThrLysSerLysGlnTyrAsnSerLysValAspGluLeu 44
UY 173 GACACGGCTGTCACATCGGCAAGAAATCTAAATTTGTGATGCTTCACATCTGCAAC 232
DB 45 SerAlaGluLeuIleTrpGluLysAsnLeuLysHisLysLeuHisAsnLeuGluAla 64
UY 233 TCAATGGCAATGCAATTTATGATGATTTAGATATGAACTATGAGGATATGATG 292
DB 65 SerLeuGluValHisThrTrpGluLeuAlaMetAsnHisLeuGluLysPheMetThrSerGlu 84
UY 293 GAATGATATTTTATGATGTTTCTTATGAGCTTCCACGCAATGCGAGCAATGTCAC 352
DB 85 GluValValGluLysMetThrGluLeuLysValProProSer----- 98
UY 353 IATAGCTAACTTAATCACAATTC-----CTGAT 385
DB 99 -- ArgSerHisSerAsnAspThrLeuLysIleProAspTrpGluLysArgThrProAsp 117
UY 386 TGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 445
DB 118 SerIleAspTrpArgLysLysGlyValValIleThrProValLysAsnGlnGlyGlnGly 137
UY 446 GTTGTGATGATTTTATGATGTTTCTTATGAGCTTCCACGCAATGCGAGCAATG 505
DB 138 SerCysTrpAlaPheSerSerValGlyValLeuGlnLysGlnLeuLysLysLysThrGly 157
UY 506 AACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 565
DB 158 LysLeuLeuAsnLeuSerProGlnAsnLeuValAspCysValSerGluAsnTyr----- 175
UY 566 AAGAGCTGAATGAGTTTATGATGAACTGTTTCAATGATTTATGATGATGATGAG 625
DB 176 --- GlyCysGlyGlyGlyTyrMetThrAsnAlaIleGlnTyrValGlnIleAsnArgGly 194
UY 626 ATTATTTAGAAAGTTTCTATGCTTACAAAGCATGAATGAGAGAGAGAGAGAGAG 685
DB 195 IleAspSerGluAsnAlaLysProTyrValGlyGlnAspGluSerCysMetTyrAsnPro 214
UY 686 AAAAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 745
DB 235 ThrGlyLysAlaIleLysCysAsnGlyTyrIleArgGluIleProGlnIleAsnGluLysAla 234
  
```

UY 746 TTAAGAGAGCTGAGGCTAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
 DB 235 LeuLysArgAlaValAlaArgValGlyProValSerValAlaIleAspAlaSerLeuThr 254
 UY 806 TCTTCTCCCTACACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862
 DB 255 SerPheGlnPheTyrSerLysGlyValTyrTyrAspGluAsnGlySerSerAsnVal 274
 UY 863 AATCATGAGATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 922
 DB 275 AsnHisAlaValLeuAlaValGlyTyrGlyIleGlnLysGlyAsnLysHisTrpIleGle 294
 UY 923 AAAACAGCTGGGGCTCAACTTGTGTAACCAAGATATATGATGAGAGAGAGAG 982
 DB 295 LysAsnSerTrpGlyGluSerTrpGlyAsnLysGlyTyrIleLeuMetAlaArgAsnLys 314
 UY 983 GGAATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
 DB 315 AsnAlaLysGlyIleAlaAsnLeuAlaSerPheProLysMet 329

RESULT 10
 US-08-330-121B-3
 : Sequence 3, Application US/08330121B
 : Patent No. 5736357
 : GENERAL INFORMATION:
 : APPLICANT: Bromme, Dieter
 : APPLICANT: Okamoto, Kathleen
 : TITLE OF INVENTION: CATHEPSIN O2 PROPHASE
 : NUMBER OF SEQUENCES: 11
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Flehr, Hohbach, Test, Albritton
 : ADDRESS: 6 Herbert
 : STREET: Four Embarcadero Center, Suite 3400
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: United States
 : ZIP: 94111-4187
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/330,121B
 : FILING DATE: 27-OCT-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Silva, Robin M.
 : REGISTRATION NUMBER: 38,304
 : REFERENCE/DOCKET NUMBER: A-60261/DJB/RMS
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 781-1989
 : TELEFAX: (415) 398-3249
 : TELEX: 910 277299
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 329 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: unknown
 : TOPOLOGY: unknown
 : MOLECULE TYPE: protein
 : US-08-330-121B-3

Alignment Scores:
 Pred. No.: 5,260-97
 Score: 944.50
 Percent Similarity: 69.55%
 Best Local Similarity: 55.52%
 Query Match: 47.10%
 Matches: 186
 Conservative: 47
 Mismatches: 79
 Indels: 24
 Gaps: 5

US-10-010-580-1 (1-1126) x US-08-330-121B-3 (1-329)


```

QY 11 CATGTGAATCTTGGTAAGAAAGAAACCTACACCAACCAATCAAGCAAGACAT -GAGGAAAGTA 17
Db 25 GlnTrpGluLeuTrpLysLysThrLysSerLysGlnTrpAsnSerLysValAspGluIle 44
QY 173 GCACGGCGTCTCATCTCGGCAAAAAAACTAAATTTGGATGCTCTCACAATCTGGAACAC 23
Db 45 SerArgArgLeuLeuIleTrpGluLysAsnLeuLysHisIleSerIleHisAsnLeuGluAla 64
QY 233 TCATGGGAATGCCATTCATCATATGATCAAGCAATGAAACCAATCGCGGACATGCGTCAA 29
Db 65 SerLeuLysAlaHisThrLysGluLeuAlaMetAsnHisLeuGluLysMetHisSerLys 84
QY 293 GAAATGATAATCTTTGATGAGATTCGCTGAGAGATTTCCAGCCAAATGGCAGCAAAATGTCACT 35
Db 85 GluValValGlnLysMetThrGlyLeuLysValProProSer - - - - - 98
QY 353 TATAGCTCAACCTCATATCAAGAAATG - - - - - CCGCAT 38
Db 99 - - - ArgSerHisSerAsnAspThrLeuLysrIleProAspTrpGluLysArgThrProAsp 117
QY 386 TCTGTGTGACAGGAGACAGCAAGCGGCTGTGTACTGCAAGTGAAGCAAAATACACAGGTTCTTGCGT 44
Db 118 SerLLeaspLysrArgLysGlyLysGlyValThrProValLysAsnGlnGlyGlnCysGly 137
QY 446 GCTTTGTGGCTTTCACAGCTGTGGGGGCGCTGGAAAGCAACAATCAAGCTAAAAACAGA 50
Db 138 SerCysTrpAlaPheSerSerValGlyAlaLeuGlnGluGlnLeuLysLysLysThrCly 157
QY 506 AACCTGCTGTCTGTGATGTCAGACGAAACCTTGATAGATTCGTCAACGAAAAATATGCGCAT 56
Db 158 LysLeuLeuAsnLeuSerProGlnAsnLeuValAspCysValSerGluAsnTrp - - - - - 175
QY 566 AAAGCGTCGAATGGCGGCTTTCATAGACACACTGCTTTCGAGTATATTATGATTAACAAGCGC 62
Db 176 - - - GlySerGlyGlyGlyLysMetThrAsnAlaPheGlnTrpValGlnArgAsnArgLys 194
QY 626 ATTGATCAAAACCTTCCTATCCCTCTCAAAAGCCATGATGAAAGTGCAGATATGACTCA 68
Db 195 IleAspSerLysAspAlaTrpProLysValGlyGlnAspGluSerCysMetLysAsnPro 214
QY 686 AAAAAGCGACCTGCCACATCTTCAAAAGTATACTGCAACTCTCCCTTGGCGAGCAAGATGGC 74
Db 215 ThrGlyLysAlaLalaLysCysAsnGlyLysArgGluIleProGluGluLysAsnGlnLysAla 233
QY 746 TTTAAAGAACCTGtGGCCAAIAAAGAACCTGtGtGtGtGtGtGtATATAGATGACACCCACTAT 80
Db 235 LeuLysArgAlaValAlaAlaArgValGlyProValSerValAlaAlaIleAspAlaSerLeuThr 254
QY 806 TCTTTCTCTCTGACAGAAAGTGGTGTCTCTATGAAACCATGCTGT - - - ACACAGAAATGG 86
Db 255 SerPheGlnPheLysSerLysGlyValLysTrpAspGluAsnCysSerSerAspAsnAla 274
QY 863 AATCATGCACTATTACTGCTGTGGCTATGCTAACTTAACTGGCAAAAGCACTAGCTGCTGTG 92
Db 275 AsnHisAlaValLeuAlaValGlyLysGlyIleGlnLysGluLysAsnLysHisTrpLeuIle 294
QY 923 AAAAAGAGCGGGGGCGCTCAACTTTGGTGAACAGAGATATATCGCATGGCAAAAGACGT 98
Db 295 LysAsnSerTrpLysLysLysTrpGlyAsnLysGlyLysLysLeuMetAlaAlaArgAsnLys 314
QY 983 GTAATCATATGTAATTAAGTATATCTTATATCTTATCAAAATG 1027
Db 315 AsnAsnAlaCysGlyIleAlaAsnLeuAlaSerPheProLysMet 329

```

```

1 STREET: Four Embarcadero Center, Suite 3400
2 City: San Francisco
3 State: California
4 Country: United States
5 Zip: 94111-4187
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: Patentin Release #1.0, Version #1.30
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: PCT/US95/13820
13 FILING DATE: 26-OCT-1995
14 CLASSIFICATION:
15 APPLICATION NUMBER: US UNKNOWN
16 FILING DATE: 02-OCT-1995
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: US 08/330,121
19 FILING DATE: 27-OCT-1994
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Silva, Robin M.
22 REGISTRATION NUMBER: 38,304
23 REFERENCE/DOCKET NUMBER: FP-60261-1-PC/DJ/B/RMS
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: (415) 781-1989
26 TELEFAX: (415) 398-3249
27 TELEX: 910 277299
28 INFORMATION FOR SEQ. ID NO: 3:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 329 amino acids
31 TYPE: amino acid
32 STRANDEDNESS: unknown
33 TOPOLOGY: unknown
34 MOLECULE TYPE: protein
35 PCT-US95-13820-3
36
37 Alignment Scores:
38 Pred. No.: 5,260-97 Length: 329
39 Score: 944.50 Matches: 186
40 Percent Similarity: 69.55% Conservative: 47
41 Best Local Similarity: 55.52% Mismatches: 79
42 Query Match: 47,308 Indels: 24
43 Gaps: 5
44 DB:
45
46 US-10-010-580-1 (1-1126) x PCT-US95-13820-3 (1-329)
47
48 Oy 54 CTGTTCCGCTTGGTCCCTATGCGAGGCGACAAAGTACATAAAGTCCGCTGGATCAT 113
49 |||||||..... |||||||..... |||||||..... |||||||.....
50 Db 8 LeuIeuProValAlValSerPheAla-----LeuHisProGluGluIleLeuAlaSPThr 24
51
52 Oy 114 CACTGGAATCTCTGGAAGAAACCTACAGCAACAACATCAGCAAGACGAAT 172
53 |||||..... |||||||..... |||||||..... |||||||.....
54 Db 25 GluIleProIuIeuIeuIleProGluIuIysAluIeuIysHisIleSerIleHisAluIeuGluAla 64
55
56 Oy 173 GCACGGGGTCTCTCTGGGAAAAAATCAAAATTGTGATCTTGGATCAATCTGGAAAT 243
57 ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
58 Db 45 SerArgIleGluIleIleProGluIuIysAluIeuIysHisIleSerIleHisAluIeuGluAla 64
59
60 Oy 233 TTAATGGGAATGATCATCATATATATATATATATATATATATATATATATATATATAT 292
61 |||||||..... |||||||..... |||||||..... |||||||.....
62 Db 65 SerIeuGluIuIhisIleThrIleGluIeuAlaIleAlaHisIleGluIuIysAlaSPMetThrSerGlu 84
63
64 Oy 293 GAAGTATATCTTTCATGGCTTCCCTGAGAGCTTCCCAACCAATGGCAGACAAATGTCAC 352
65 |||||||..... |||||||..... |||||||..... |||||||.....
66 Db 85 GluValValGluIuIysMetThrGluIeuIysValIleProPheSer----- 98
67
68 Oy 353 TATAGTCAAACTCTATACAGAAATGG-----CTGGAT 385
69 |||||||..... |||||||..... |||||||..... |||||||.....
70 Db 99 ---ArgSerHisSerIuIysAluIeuIysHisIleSerIleHisAluIeuGluAla 64
71
72 Oy 386 TCTGTGGATGGAGAGAAAGAGATGTGTGTACTGAAATGGAAATACAGGATTTCTTGGGT 445
73 |||||||..... |||||||..... |||||||..... |||||||.....

```



```
OY 827 GGTCCTACTATGAACCATCTGT--ACTCAGAATGTGAATCATGGAGTATTAGTGT 883
    |||||
Db 262 C|YValIYrTYrAspG|uSerCysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 281
    |||||
OY 884 GCCTATGCTAACCTTATATGGCAAGACTACTGCTGTGTGAAAAACAGCTGGGGCTCAAC 943
    |||||
Db 282 G|YTYrG|Yl|eG|nLysG|YAsnLysHisTrp|leI|eLysAsnSerTrpG|YGl|uAsn 301
    |||||
OY 944 TTTCGTGACCAAGCATATATTCGATGGCAAGAAACAGTGCATATCAGTGTGGATTGCT 1003
    |||||
Db 302 TrpG|YAsnLysG|YTYr|leuMetAlaArgAsnLysAsnAsnAlaCysG|Yl|GAla 321
    |||||
OY 1004 AGTTATCCCTCTTACCCAGAAATC 1027
    |||||
Db 322 AsnLeuAlaSerPheProLysMet 329
    |||||
```

Search completed: June 27, 2003, 11:35:47
Job time : 29.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:33:19 ; Search time 57.5 Seconds

(Without alignments)
4294.342 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1997
Sequence: 1 gcatattagttatggtatgaaatgaatgtctatgacat 1126

Scoring table:

BLASTSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Dgapop 6.0, Dgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 424699 seqs, 106646833 residues

Total number of hits satisfying chosen parameters: 849398

Minimum hit seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xip
-O=/cgn2_6/USPRO.spool/US10010580/runat_27062003_104539_11849/app_query.fasta_1.1287
-DB=Published Applications AA -OPMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=local -OUTFMT=fig -NCRM=ext -HEAPSIZE=500 -MINLEN=0
-MAJEN=2000000000 -USER=0510010580 -CGN_1_1_17 -runat_27062003_104539_11849
-NCPU=6 -ICPU=3 -NM=MMAP -LARGEDUERY -NEG_SCORES=0 -WAIT -DSBLCK=100
-LONCLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YCAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2:	/cgn2_6/ptodata/1/pubpaa/FCF_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB.pep:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1775	88.9	331	9	US-10-010-580-2
2	1569	78.6	331	9	US-10-009-275-2
3	1565	78.4	331	9	US-09-990-064-2
4	1562	78.2	331	10	US-09-953-956-8

	5	1562	78.2	331	12	US-10-114-464-8	Sequence 8, Appli
	6	1545.5	77.4	330	10	US-10-010-577-2	Sequence 2, Appli
	7	944.5	47.3	329	10	US-09-953-956-7	Sequence 7, Appli
	8	944.5	47.3	329	12	US-10-114-464-7	Sequence 7, Appli
	9	939.5	47.0	329	10	US-09-953-956-2	Sequence 2, Appli
	10	939.5	47.0	329	12	US-10-114-464-2	Sequence 2, Appli
	11	939.5	47.0	361	10	US-09-925-300-1693	Sequence 1693, Ap
	12	843	42.2	334	9	US-10-028-072-12	Sequence 12, Appl
	13	843	42.2	334	9	US-10-121-049-12	Sequence 12, Appl
	14	843	42.2	334	9	US-10-123-904-12	Sequence 12, Appl
	15	843	42.2	334	9	US-10-140-470-12	Sequence 12, Appl
	16	843	42.2	334	9	US-10-175-746-12	Sequence 12, Appl
	17	843	42.2	334	9	US-10-176-918-12	Sequence 12, Appl
	18	843	42.2	334	9	US-10-176-925-12	Sequence 12, Appl
	19	843	42.2	334	9	US-10-137-865-12	Sequence 12, Appl
	20	843	42.2	334	9	US-10-140-474-12	Sequence 12, Appl
	21	843	42.2	334	9	US-10-142-431-12	Sequence 12, Appl
	22	843	42.2	334	9	US-10-143-114-12	Sequence 12, Appl
	23	843	42.2	334	9	US-10-140-002-12	Sequence 12, Appl
	24	843	42.2	334	9	US-10-142-419-12	Sequence 12, Appl
	25	843	42.2	334	9	US-10-123-262-12	Sequence 12, Appl
	26	843	42.2	334	9	US-10-142-423-12	Sequence 12, Appl
	27	843	42.2	334	9	US-10-121-050-12	Sequence 12, Appl
	28	843	42.2	334	9	US-10-141-755-12	Sequence 12, Appl
	29	843	42.2	334	9	US-10-143-032-12	Sequence 12, Appl
	30	843	42.2	334	9	US-10-123-108-12	Sequence 12, Appl
	31	843	42.2	334	9	US-10-123-236-12	Sequence 12, Appl
	32	843	42.2	334	9	US-10-123-261-12	Sequence 12, Appl
	33	843	42.2	334	9	US-10-140-921-12	Sequence 12, Appl
	34	843	42.2	334	9	US-10-140-928-12	Sequence 12, Appl
	35	843	42.2	334	9	US-10-121-045-12	Sequence 12, Appl
	36	843	42.2	334	9	US-10-123-292-12	Sequence 12, Appl
	37	843	42.2	334	9	US-10-123-903-12	Sequence 12, Appl
	38	843	42.2	334	9	US-10-124-819-12	Sequence 12, Appl
	39	843	42.2	334	9	US-10-124-822-12	Sequence 12, Appl
	40	843	42.2	334	9	US-10-140-925-12	Sequence 12, Appl
	41	843	42.2	334	9	US-10-160-498-12	Sequence 12, Appl
	42	843	42.2	334	9	US-10-121-041-12	Sequence 12, Appl
	43	843	42.2	334	9	US-10-121-043-12	Sequence 12, Appl
	44	843	42.2	334	9	US-10-121-047-12	Sequence 12, Appl
	45	843	42.2	334	9	US-10-123-215-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-010-580-2
; Sequence 2, Application US/10010580
; Publication No. US20030104972A1
; GENERAL INFORMATION:
; APPLICANT: Unimond, robin 1
; APPLICANT: Baker, sherry
; APPLICANT: Karlsson, lars
; TITLE OF INVENTION: polynucleotide and polypeptide sequences of canine cathepsin s
; FILE REFERENCE: ORT-1458
; CURRENT APPLICATION NUMBER: US/10/010,580
; CURRENT FILING DATE: 2001-11-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: canine
US-10-010-580-2

Alignment Scores:

Pred. No.:	8,096-180	Length:	331
Score:	1775.00	Matches:	330
Percent Similarity:	99.70%	Conservative:	0
Best Local Similarity:	99.70%	Mismatches:	1
Query Match:	88.88%	Indels:	1
DB:	9	Gaps:	0


```
Alignment Scores:
Pred. No.: 9.99e-81
Score: 843.00
Percent Similarity: 65.368
Best Local Similarity: 50.608
Query Match: 42.218
DB: 9
```

US-10-010-580-1 (1-1126) x US-10-028-072-12 (1-334)

QY	5	CTGCTTCGCGTGGTGGCTCCATATGAGTGGCAATAGATACATAAA---GATCTTCACCTGGAT	110
Db	7	LeuAlaAlaIphCysLeuClYlIleAlaSerAlaValProIlysPheAspCInAsnLeuAsp	26
QY	111	CATCATGGATATCTCTGGAAAGAAAACCTACAGCAACCAATACAGGAAGACATATGGAG	170
Db	27	ThrIlystrPtyrClnIrrIrrPylsalIathrlIsArqArq---LeuIyrGlyAlAlasnGlnIuG	46
QY	171	TACACGGCGCTCTCATCTCGGAAAAAAATCTIAAAATTTGTATGCTTCACAACTCGAAC	230
Db	46	IYrIrrPArqArqAlaValIrrPqIuIlySasMetIlySmetIleGlnIleuHisAsnGlnIyr	66
QY	231	ACTCATGGGAATGATCATATCATATGATCTGTGGCATACATCTGGAGCAATAGCTAGTG	290
Db	66	YIrrClnGlnIlyShIlyShIlyCylPheIrrMetAlaIleAsnAlaIphClnIlyAspMetIrrAsnG	86
QY	291	MAGAAGTATATCTTGATGGATGTTTCTTCAATATTCAGAAATATCTCA	350
Db	86	IuGlnIphArqGlnIleMetIlgIlyCysPheArq---AsnGlnIlyPheArqIlySylIySv	105
QY	351	CTTATAGCTCAAACTCTTAATCAGAAATTCCTGATTTCTGTGACTGCAGACAGAGGGCT	410
Db	105	AlpheArqGlnProlLeuPheIleuAsnIleuAsnProIlySsrValAspIrrPArqIlySylIyT	125
QY	411	GTTTATCTAGTGAATATATAGGGTCTTCTTGGTCTTCTTGGATTTTATAGTCTGA	470
Db	125	YrAlIrrIrrProValIlySasGlnIlySasGlnIlySasGlnIySsrCysIrrPAlaIphSerAlaIrrC	145
QY	471	GGGCCCTGGAAAGCAATATGAAGATATAAAATAGAAATATATATTTATATATATATAT	530
Db	145	IlyAlaIleuGlnIlyGlnIleMetPheArqIlySthrGlyIlyLeuValSerIleuSerClnIuA	165
QY	531	ACTTGTACATTCTCACTGAAAAATATAGGAATATAAGCTGCATATCGCGTTTATCA	590
Db	165	snIleuValAspCysSerIrrArqProGln---GlyAsnGlnIlyCysAsnGlnIyGlyPheMetA	184

[illegible]

```

RESULT 13
US-10-121-049-12
: Sequence 12, Application US/10121049
: Publication No. US2003002239A1
: GENERAL INFORMATION
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvarolt, Ellen
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3330R1C17
: CURRENT APPLICATION NUMBER: US/10/121.049
: CURRENT FILING DATE: 2002-04-12
: Prior Application removed - See File Wrapper of Palm
: NUMBER OF SEQ ID NOS: 550
: SEQ ID NO 12
: LENGTH: 334
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-121-049-12

Alignment Scores:
Score: 9 99% 334
Percent Similarity: 843.00
Percent Local Similarity: 65.36%
Best Local Similarity: 50.60%
Query Match: 42.21%
Indels: 10
Length: 334
Matches: 168
Conservative: 49
Mismatch: 105
Indels: 10

```

DB:	9	Gaps:	5
US-10-010-580-1 (1-1126) X US-10-121-049-12 (1-334)			

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:28:49 ; Search time 286.5 Seconds

(without alignments)
5067,850 Million cell updates/sec

Title: US-10-010-580-1

1997

Perfect score: 1 gcatattatgaatgagagca.....ataaatgtgcatgacacat 1126

Sequence:

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_n2p_model -DEV=xip
-DB=/cgn2_1/USFtu.shou/US100/0580/runat_27062003.104538.11797/4pp_guery.fasta_1.1287
-DB-Pending_Patents_AA_Main -QFMT=fastan -SUFFIX=n2p_rpm -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -ONITS=hits -START=1 -END=1 -MATRIX=blomsu62
-TRANS=human40 -31 -LIST=45 -D=ALIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USFR=US100/0580 -ECGN_1_1_5b2 -runat_27062003.104538.11797
-NCPU=6 -ICPU=3 -NO_MMAP -LANG=QUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONSLUG -DEV TIMEOUT=120 -WARM TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

- 1: Pending_Patents_AA_Main.*
- 2: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US06.COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US07.COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
- 27: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1775	88.9	331	24	US-10-010-580-2	Sequence 2, Appli
2	1581.5	79.2	353	21	US-09-760-475-1771	Sequence 1771, Ap
3	1581.5	79.2	353	21	US-09-760-475-2997	Sequence 2997, Ap
4	1571	78.7	331	21	US-09-791-537-5974	Sequence 5974, Ap
5	1571	78.7	827	27	US-60-230-435-1540	Sequence 1540, Ap
6	1569	78.6	331	9	US-08-553-125-8	Sequence 8, Appli
7	1569	78.6	331	16	US-09-219-441-8	Sequence 8, Appli
8	1569	78.6	331	21	US-09-791-537-94624	Sequence 94624, A
9	1569	78.6	331	24	US-10-099-275-2	Sequence 2, Appli
10	1565	78.4	331	23	US-09-990-064-2	Sequence 2, Appli
11	1562	78.2	331	9	US-08-553-125A-8	Sequence 8, Appli
12	1562	78.2	331	16	US-09-219-441A-8	Sequence 8, Appli
13	1562	78.2	331	18	US-09-463-000-2	Sequence 2, Appli
14	1562	78.2	331	21	US-09-791-537-90091	Sequence 90091, A
15	1562	78.2	331	22	US-09-888-622-7	Sequence 2, Appli
16	1562	78.2	331	23	US-09-953-956-8	Sequence 8, Appli
17	1562	78.2	331	25	US-10-114-401-2	Sequence 2, Appli
18	1562	78.2	331	25	US-10-114-454-8	Sequence 8, Appli
19	1560	78.1	331	9	US-08-553-861-4	Sequence 4, Appli
20	1545.5	77.4	330	24	US-10-010-577-2	Sequence 2, Appli
21	1543	77.3	330	26	US-10-321-279-11449	Sequence 11449, A
22	1513.5	75.8	355	1	PCT-US01-14827-13870	Sequence 14870, A
23	1373.5	68.8	340	21	US-09-791-537-96383	Sequence 96383, A
24	1367.5	68.5	340	21	US-09-791-537-32061	Sequence 32061, A
25	1338.5	67.0	330	21	US-09-791-537-13081	Sequence 13081, A
26	1334.5	66.8	330	21	US-09-791-537-120906	Sequence 120906, A
27	1328	61.5	267	27	US-60-389-987-1343	Sequence 1343, Ap
28	1228	61.5	267	27	US-60-412-418-1343	Sequence 1343, Ap
29	1228	61.5	268	21	US-09-791-537-152232	Sequence 152232, A
30	1089	54.5	217	21	US-09-791-537-16900	Sequence 16900, A
31	1016.5	50.9	331	21	US-09-791-537-1024	Sequence 1024, Ap
32	978	49.0	196	21	US-09-791-537-5111	Sequence 5111, Ap
33	944.5	47.3	329	9	US-08-553-861-3	Sequence 3, Appli
34	944.5	47.3	329	9	US-08-553-125-7	Sequence 7, Appli
35	944.5	47.3	329	9	US-08-553-125A-7	Sequence 7, Appli
36	944.5	47.3	329	16	US-09-219-441-7	Sequence 7, Appli
37	944.5	47.3	329	16	US-09-219-441A-7	Sequence 7, Appli
38	944.5	47.3	329	21	US-09-791-537-10429	Sequence 10429, A
39	944.5	47.3	329	23	US-09-953-956-7	Sequence 7, Appli
40	944.5	47.3	329	25	US-10-114-464-7	Sequence 7, Appli
41	943.5	47.2	329	21	US-09-791-537-104875	Sequence 104875, A
42	940.5	47.1	329	19	US-09-548-512-2	Sequence 2, Appli
43	940.5	47.1	329	19	US-09-548-512-2	Sequence 2, Appli
44	940.5	47.1	329	24	US-10-016-333-4	Sequence 4, Appli
45	939.5	47.0	329	1	PCT-US96-10346-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-10-010-580-2
: Sequence 2, Application US/10010580
: GENERAL INFORMATION:
: APPLICANT: thurmond, robin 1
: APPLICANT: baker, sherry
: APPLICANT: karlsson, lars
: TITLE OF INVENTION: polynucleotide and polypeptide sequences of canine cathepsin s
: FILE REFERENCE: OPT-1458
: CURRENT APPLICATION NUMBER: us/10/010.580
: CURRENT FILING DATE: 2001-11-08
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 331


```
QY 542 TGCTCAACTGAAAAATATCGCAATAAAGCGTGCATGCGGCTTTCATGACAACTGCTTTC 601
DB 192 CysSerThrGluValGlyValAsnLysGlyCysAsnGlyGlyPheMetThrThrAlaPhe 211
QY 602 CAGTATATATTGATTAACACGCGATTGATTGAGAGCTTCTATCCCTCAAAAGCAGT 661
DB 212 GlnTyrIleIleAspAsnLysGlyIleAspSerAspAlaSerTyrProTyrLysAlaMet 231
QY 662 AATGAAAGTGCATATATGACTCAAAAAGCGAGCTGCCATGCTGCAAGTATAGTAA 721
DB 232 AspGlnLysCysGlnTyrAspSerLysTyrArgAlaIleThrCysSerLysTyrThrGlu 251
QY 722 CTTCCTTTCCGACGATGACATGCGCTTAAAGACAGCTGCGCAATTAAGACCTGCTCT 781
DB 252 LeuProTyrGlyArgGlyAspValLeuLysGlyIleValAlaAsnLysGlyProValSer 271
QY 782 GTGGTATAGATGCGAGCCACTATCTTCTCTCTGACAGAGTGGTGTACTATGAA 841
DB 272 ValGlyValAspAlaIleArgHisProSerPheLeuTyrArgSerGlyValTyrGlu 291
QY 842 CCATCCCTGACTCAAGATGATGATCATGATATATGATGCTGCTATGATACCTTAAT 901
DB 292 ProSerCysThrGlnAsnValAsnHisGlyValLeuValValGlyTyrGlyAspLeuAsn 311
QY 902 GCGAAGACTACTGCGCTTGTGAAAAACAGCTGGGCGCTCAACTTGTGACCAAGATAT 961
DB 312 GlyLysGluTyrTrpLeuValLysAsnSerTrpGlyHisAsnPheGlyGluGlyTyr 331
QY 962 ATTGGGATGCGCAAGAAACAGTGGAAATCTGTGGATGCTACTTATCCCTCTACCCA 1021
DB 332 IleArgMetAlaIleArgAsnLysGlyAsnHisCysGlyIleAlaSerPheProSerTyrPro 351
QY 1022 GAAATC 1027
DB 352 GluIle 353

RESULT 3
US-09-760-475-2997
: Sequence 2997, Application US/09760475
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P749
: CURRENT APPLICATION NUMBER: US/09/760,475
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 4122
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2997
: LENGTH: 353
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-760-475-2997

Alignment Scores:
Pred. No.: 1,08e-158 Length: 353
Score: 1581.50 Matches: 295
Percent Similarity: 92.69% Conservative: 22
Best Local Similarity: 86.26% Mismatches: 24
Query Match: 79.19% Indels: 2
DB: 21 Gaps: 1

US-10-010-580-1 (1-1126) x US-09-760-475-2997 (1-353)
```

```
DB 52 LeuTrpLysLysThrTyrGlyLysGlnTyrLysGlyIleValAsnIleValAlaArgArg 71
QY 182 CTGATCTGGGAAAAAATCTAAATTTGTGATGCTTTCACAAATGGAACACTGAATGGA 241
DB 72 LeuIleTrpLysLysAsnLysLysPheValIleMetLeuHisAsnLysIleHisSerMetCyl 91
QY 242 ATGCATTCATATGATGATAGGATGAGCAATCTGGGAGACATGACATGCTGCAAGAGTGA 301
DB 92 MetHisSerTyrAspLeuGlyMetLeuAsnHisLeuGlyAspMetThrSerGluIleValMet 111
QY 302 TCTTTGATGGCTTCCCTGAGAGTTCCAGCAATGCGACAGAAATGCTCACTTAATAGCTCA 361
DB 112 SerLeuMetSerSerLeuArgValProSerGlnTrpGlnArgAsnIleThrTyrLysSer 131
QY 362 AACTATATACGAAATTTGCTGATGTTGTGGACTGGAGAGAGAGAGGCTGTGATGAA 421
DB 132 AsnProAsnArgIleLeuProAspSerValAspTrpArgGluLysGlyCysValThrGlu 151
QY 422 GTGAAATACAGCGCTTCTGTGCGCTTGTGGCTTTCAGCGCTTGGCGGCGCTGCAA 481
DB 152 ValLysTyrGlnLysSerCysGlyAlaCysTrpAlaPheSerAlaValGlyAlaLeuGlu 171
QY 482 GCACAACTGAAGCTAAACAGAGAAAGCTGGTGTCTGTGATGCGACAGAACTTGTGTAAT 541
DB 172 AlaGlnLeuLysLysLysThrGlyLysLeuValSerLeuSerAlaGlnAsnLeuValAsp 191
QY 542 TGCTCAACTGAAAAATATGGAATTAAGCGTGAATGGCGGCTTTCATGCAACTGCTTTC 601
DB 192 CysSerThrGluLysTyrGlyAsnLysGlyCysAsnGlyCylPheMetThrThrAlaPhe 211
QY 602 CAGTATATATTGATTAACACGCGATTGATTGAGAGCTTCTATCCCTCAAAAGCAGT 661
DB 212 GlnTyrIleIleAspAsnLysGlyIleAspSerAspAlaSerTyrProTyrLysAlaMet 231
QY 662 AATGAAAGTGCATATGACTCAAAAAGCGAGCTGCCATGCTGCAAGTATAGTAA 721
DB 232 AspGlnLysCysGlnTyrAspSerLysTyrArgAlaIleThrCysSerLysTyrThrGlu 251
QY 722 CTTCCTTTCCGACGATGACATGCGCTTAAAGACAGCTGCGCAATTAAGACCTGCTCT 781
DB 252 LeuProTyrGlyArgGlyAspValLeuLysGlyIleValAlaAsnLysGlyProValSer 271
QY 782 GTGGTATAGATGCGAGCCACTATCTTCTCTCTGACAGAGTGGTGTACTATGAA 841
DB 272 ValGlyValAspAlaIleArgHisProSerPheLeuTyrArgSerGlyValTyrGlu 291
QY 842 CCATCCCTGACTCAAGATGATGATCATGATATATGATGCTGCTATGATACCTTAAT 901
DB 292 ProSerCysThrGlnAsnValAsnHisGlyValLeuValValGlyTyrGlyAspLeuAsn 311
QY 902 GCGAAGACTACTGCGCTTGTGAAAAACAGCTGGGCGCTCAACTTGTGACCAAGATAT 961
DB 312 GlyLysGluTyrTrpLeuValLysAsnSerTrpGlyHisAsnPheGlyGluGlyTyr 331
QY 962 ATTGGGATGCGCAAGAAACAGCTGGAATCTGTGGATGCTACTTATCCCTCTACCCA 1021
DB 332 IleArgMetAlaIleArgAsnLysGlyAsnHisCysGlyIleAlaSerPheProSerTyrPro 351
QY 1022 GAAATC 1027
DB 352 GluIle 353

RESULT 4
US-09-791-537-5974
: Sequence 5974, Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Biomimix, Inc.
: APPLICANT: Debe, Derek
: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
: FILE REFERENCE: 261/210
```



```

|||||
Db 637 AlapheserAlaValGlyAlaLeuAlaGlnLeuLysLeuLysThrGlyLysLeuVal 656
QY 515 TCTCTGACTGCAAAATTTGGTAGATTGCTCAACTGAAAATATGGAAATAAAGGCTGC 574
Db 657 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlyLysGlyAsnLysGlyCys 676
QY 575 AATGGCGTTTCATGACAACTGCTTCAGATATATTATGATTAACAAACGGCATGATCA 634
Db 677 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleIleAspAsnLysGlyIleAspSer 696
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGTGCAGATATGACTCAAAAAACGCA 694
Db 697 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 716
QY 695 GGTGGCCATGTTTAAAGTATATGCAACTGCTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
Db 717 AlaAlaThrCysSerLysTyrThrGlyLeuProTyrGlyArgGlnAspValLeuLysGln 736
QY 755 GCTGTGGCCAAATAAAGACCTGTGTCTGTGGCTATAGATGGACCCACTATTTCTTCTTC 814
Db 737 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 756
QY 815 CTGTACAGAACTGGTGTCTACTATGACACCATCTGTACTCGAATGTGAATCATGAGATA 874
Db 757 LeuTyrArgSerGlyValTyrTyrGlyProSerCysThrGlnAsnValAsnHisGlyVal 776
QY 875 TTAGGTGTGGCTATGTGAACCTTAATGGAAAGCTGCTGGTGTGAAGAAACAGCGCG 934
Db 777 LeuValValGlyTyrGlyAspLeuAsnLysGlnTyrTyrPheValLysAsnSerTyr 796
QY 935 GCGCTCAACTTGTGTGACCAAGATATATTCGATGGCAGAAACAGTGGAAATCACTGT 994
Db 797 GlyHisAsnPheGlyGlnGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisGly 816
QY 995 GGGATTGCTACTTATCCCTCTTACCCAGAAATC 1027
Db 817 GlyIleAlaSerPheProSerTyrProGluLeu 827

RESULT 6
US-08-553-125-8
: Sequence 8, Application US/08553125
: GENERAL INFORMATION:
: APPLICANT: HASTINGS, ET AL.
: TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: STREET: 6 BRCKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/553,125
: FILING DATE: Herewith
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/208,007
: FILING DATE: March 8, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: FERRARO, GREGORY D.
: REGISTRATION NUMBER: 36,134
: REFERENCE/DOCKET NUMBER: 325800-505
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744

```

```

: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 331 AMINO ACIDS
: TYPE: AMINO ACID
: STRANDEDNESS:
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
: US-08-553-125-8

Alignment Scores:
Pred. No.: 2,272-157 Length: 331
Score: 1569.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.57% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-08-553-125-8 (1-331)

QY 36 ATGAATGCTAGTGGGCTGCTTCCGCTTATGCACTGGCAGCAATGAATA 95
Db 1 MetLysArgLeuValCysValLeuLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
QY 96 GATCCCACTGTGATCATCACTGAAATCTCTGCAAGAAAACCTACAGCAAAATC- AAG 154
Db 21 AspProThrLeuAspHisHisIleThrHisLeuTyrLysThrTyrGlyLysGlnTyrLys 40
QY 155 GAAGAGATAGAGAAAGACGACGCGCTCATATCTGGGAAAAAATTTAAATTTGTATG 214
Db 41 GlyLysAsnGlnGlnAlaValArgArgLeuIleProLysAsnLeuLysPheValMet 60
QY 215 CTTCACAAATGTGAACACCTCAATGGCAATGATCATATGATGTAGCATGAACATCTG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuIleMetAsnHisLeu 80
QY 275 GGAGACATGACCTGTGGAAGAAGTATATCTTATGAGGTTCCTGACAGCTTCCAGCCAA 334
Db 81 GlyAspMetThrSerGlnGlnValMetSerLeuMetSerSerLeuArgValProSerGln 100
QY 335 TGGCAGAGAAATGTCACCTTATAGTCAAACTTAATCATGAATATGGCTCATTTGTGAC 394
Db 101 TrpGlnArgAsnIleThrTyrTyrLysSerAspProAsnArgIleLeuProAspSerValAsp 120
QY 395 TGGAGAGAGAAAGCGCTGTGTACTGAAGTGAATATACCAGCGTCTTGTGCTGCTTTCG 454
Db 121 TrpArgGlnLysGlnCysValThrGlnValLysTyrGlnGlySerGlnLysValAcylStrp 140
QY 455 GCTTTCAGCGCTGTGGGGCCCTGGAAAGCAACCTGAAGCTAAAAAACAGAAAGCTGTCG 514
Db 141 AlapheserAlaValGlyAlaLeuGlnAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTGTGATGTCACAGAACTGGTAGATTGCTCAACTGAAAATATGGAAATAAAGCTGAC 574
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlnLysAsnLysGlyCys 180
QY 575 AATGGCGTTTCATGACAACTGCTTCAGATATATTATGATTAACAAACGGCATGATCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleIleAspAsnLysGlyIleAspSer 200
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGTGCAGATATGACTCAAAAAACGCA 694
Db 201 AspAlaSerTyrProTyrLysAlaMetAspLeuLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGGCCATGTTTAAAGTATATGCAACTGCTTCGAGTGAAGTGCAGATATGCTTAAAGAA 754
Db 221 AlaAlaThrCysSerLysTyrThrGlnLeuProTyrGlyArgGlnAspValLeuLysGln 240
QY 755 GCTGTGGCCAAATAAAGACCTGTGTCTGTGGCTATAGATGCGACGCACTATTTCTTCTTC 814
Db 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGTACAGAGCTGTGTCTACTATGAAACCATCTGTACAGAAATGTAATCATGACAGTA 874

```



```

: TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
: FILE REFERENCE: PF107D1
: CURRENT APPLICATION NUMBER: US/08/553,125A
: CURRENT FILING DATE: 1995-11-07
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-08-553-125A-8

Alignment Scores:
Pred. No.: 1,26e-156 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Best Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-08-553-125A-8 (1-331)

OY 36 ATGAATGGCTAGTGGGCTGCTCGCTGCTGCTATGCACTGGCACAAGTACATATA 95
DB 1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
OY 96 GATCCCACTGTGATCATCACTGCAATCTGTGCAAGAAAACCTACACCAAACTC-AAG 154
DB 21 AsprothriLeuAspHisHisIstRHisLeuTrpLysIstRTrpGlyLysGlnTrpLys 40
OY 155 GAACAGATGAGAAATGACGACGGCGTCTATCTGGAAAAAAATCTAAAAATTTGATG 214
DB 41 GluLysAsnGlnGluAlaValArgLeuLleTrpGlnLysAsnLeuLysPheValMet 60
OY 215 CTTCACAATCTGCAACACTCAATGCAATGATCATATGATTAAGGCATCAACATCTG 274
DB 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlnMetAsnHisLeu 80
OY 275 GGACACATGACTGCTGAAGAAGTATCTTTGATGGGCTCCGACAGATTCACGCCAA 334
DB 81 GlyAspMetThrSerGlnGluValMetSerLeuThrSerLeuArgValProSerGln 100
OY 335 TGGCAGAGAAATGTACTATTAGTCAACCTCAATGAAATGAGAAATGCTGATTCGTGAC 394
DB 101 TrpGlnArgAsnLleThrTrpLysSerAspProAsnArgLleLeuProAspSerValAsp 120
OY 395 TGCAGACAGAGCGCTGTCTACTGAAGTGAATATACAGGCTTCTGCTGCTTTGG 454
DB 121 TrpArgGlnLysGlnCysValThrGlnValLysTrpGlnGlnSerCysGlnAlaCysTrp 140
OY 455 GCTTTCAAGCGCTGTGGGGCCCTGGAAACACACTGAACCTAAAAACAGCAAACTGCTG 514
DB 141 AlaPheSerAlaValGlnAlaLeuGlnAlaGlnLeuLysLeuLysTrpGlnLysLeuVal 160
OY 515 TCTGTGAGTGCAGCAACTGTGATGATGCTCAACTGAAATATGGGAATTAAGGGTGC 574
DB 161 ThrLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTrpGlnLysLysGlnCys 180
OY 575 AATGCGGCTTTCATGACAACTGCTTTCAGATATATTAATGATAAACAAGCATGATGCA 634
DB 181 AsnGlnGlyPheMetThrThrAlaPheGlnTrpLleLeuAspAsnLysGlnLleAspSer 200
OY 635 GAAGCTTCTGCTGCTCAACAAAGCCATGATGAAGATGACATATGACTCAAAAAACGA 694
DB 201 AspAlaSerTrpTrpTrpLysAlaIleMetAspGlnLysCysGlnTrpAspSerLysTrpArg 220
OY 695 GCTGCGACATGTTCAAGTATATGAACTTCTGCTTGGAGTGAAGATGCTTAAAGAA 754
DB 221 AlaAlaIstRTrpCysSerLysTrpThrGlnLeuProTrpGlnArgGlnAspValLeuLysGln 240
OY 755 GCTGTGGCAATTAAGACCTGTGTCTGTGCTATACATGCGACCACTATCTTTCTTC 814
DB 241 AlaValAlaAsnLysGlnProValSerValGlnValAspAlaArgHisProSerPhePhe 260
```

```

OY 815 CTGTACAGAAGTGGTGTACTACTATGAAACCATCCTGACTCAGATGTGAATGTGAAGTA 874
DB 261 LeuTrpArgSerGlnValAlaTrpTrpGlnProSerCysThrGlnAsnValAsnHisGlnVal 280
OY 875 TTAGTGGTGGCTATGATGTAACCTTAATGGAAAGACTACT-33-TTGTGAAAAA-7A-TGG 934
DB 281 LeuValValGlnGlyTrpGlnLysPheAsnGlnLysGlnTrpLleTrpLeuValLysAsnSerTrp 300
OY 935 GGCCTCAACTTTGGTGTACCAAGAGATATATTGGATGGCAAGAAACAGTGAATGACTGT 994
DB 301 GlnHisAsnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
OY 995 GGAGTGTAGTATACCTCTTACCCAGAAATC 1027
DB 321 GlyIleAlaSerPheProSerTrpTrpGlnIle 331

RESULT 12
US-09-219-441A-8
: Sequence 8, Application US/09219441A
: GENERAL INFORMATION:
: APPLICANT: Hastings, et al.
: TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
: FILE REFERENCE: PF107D2D1
: CURRENT APPLICATION NUMBER: US/09/219,441A
: CURRENT FILING DATE: 1998-12-23
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-219-441A-8

Alignment Scores:
Pred. No.: 1,26e-156 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Best Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-09-219-441A-8 (1-331)

OY 36 ATGAATGGCTAGTGGGCTGCTCGCTGCTGCTATGCACTGGCACAAGTACATATA 95
DB 1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
OY 96 GATCCCACTGTGATCATCACTGCAATCTGTGCAAGAAAACCTACACCAAACTC-AAG 154
DB 21 AsprothriLeuAspHisHisIstRHisLeuTrpLysIstRTrpGlyLysGlnTrpLys 40
OY 155 GAACAGATGAGAAATGACGACGGCGTCTATCTGGAAAAAAATCTAAAAATTTGATG 214
DB 41 GluLysAsnGlnGluAlaValArgLeuLleTrpGlnLysAsnLeuLysPheValMet 60
OY 215 CTTCACAATCTGCAACACTCAATGCAATGATCATATGATTAAGGCATCAACATCTG 274
DB 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlnMetAsnHisLeu 80
OY 275 GGACACATGACTGCTGAAGAAGTATCTTTGATGGGCTCCGACAGATTCACGCCAA 334
DB 81 GlyAspMetThrSerGlnGluValMetSerLeuThrSerLeuArgValProSerGln 100
OY 335 TGGCAGAGAAATGTACTATTAGTCAACCTCAATGAAATGAGAAATGCTGATTCGTGAC 394
DB 101 TrpGlnArgAsnLleThrTrpLysSerAspProAsnArgLleLeuProAspSerValAsp 120
OY 395 TGCAGACAGAGCGCTGTCTACTGAAGTGAATATACAGGCTTCTGCTGCTTTGG 454
DB 121 TrpArgGlnLysGlnCysValThrGlnValLysTrpGlnGlnSerCysGlnAlaCysTrp 140
OY 455 GCTTTCAAGCGCTGTGGGGCTGGAAGACCAACTGAACCTAAAAACAGCAAACTGCTG 514
```

```
Db 141 AlahheserAlaValGlyAlaIleuGluAlaGlnIleuLysLeuThrGlyLysLeuVal 160
QY 515 TCTTGTAGTGTAAAGATTTGATGCTCAACTCAAAAATATGGAATTAAGGCTGC 574
Db 161 ThrLeuSerAlaGlnAsnIleuValAspCysSerThrGluLysGlyAsnLysGlyCys 180
QY 575 AATGGCGGTTTCATGACAACTGCTTTCAGATATTTATGATACCAAGCGCATTCATCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGATGCAATGCACTCAAAAAGCGA 694
Db 201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGTCATATTTTCAAAATATGCAATTTGCTTTCATGCAATGCAATGCTTAAAGAA 754
Db 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValIleuLysGlu 240
QY 755 GCTTGCGCAATTAAGACACGCTGCTGCTGCTATGATGCGACGCCATATTCCTTTC 814
Db 241 AlaValAlaAsnLysGlyProValSerValGlyAlaAspAlaArgHisProSerPhePhe 260
QY 815 CTGTACAGAACTGCTGCTACTATGAAACCATCTCTGACTCAGAAATGCAATCATGAGTA 874
Db 261 LeuTyrArgSerGlyValIlyTyrGlyProSerCysThrGlnAsnValAsnHisGlyVal 280
QY 875 TTATGCTGCTGCTATGCTAACTTAATGGAAGAAAGTACGCTGCTGGAAGAAACAGCTG 934
Db 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
QY 935 GCGCTCAACTTGTGTCACCAAGCATATATTCGATGCGCAAGAAACAGTGGAAATCACTGT 994
Db 301 GlyHisAsnPheGlyGluGlnGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
QY 995 GCGATGCTACTATATCCCTCTTACCGCAAAATC 1027
Db 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 13
US-09-463-000-2
; Sequence 2: Application US/09463000
; GENERAL INFORMATION:
; APPLICANT: Abdel Meguid, Sherin
; APPLICANT: Janson, Cheryl
; APPLICANT: Lalonde, Judith
; APPLICANT: Smith, Ward
; APPLICANT: Zhao, Baoquan
; TITLE OF INVENTION: Method of Inhibiting Cathepsin K
; FILE REFERENCE: P50684
; CURRENT APPLICATION NUMBER: US/09/463,000
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: 60/052,831
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-463-000-2

Alignment Scores:
Pred. No.: 1,26e-156 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Best Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: 18 Gaps: 0

US-10-010-580-1 (1-1126) x US-09-463-000-2 (1-331)
QY 36 ATCAAAATGGCTAGTGGCTGCTGCTGCTGCTATGCGACAGGCAAGATGATAA 95
```

```
Db 1 MetLysArgLeuValLysValIleuLeuValCysSerSerAlaValAlaGlnIleuHisLys 20
QY 96 GATGCCACCTGTCATCATCATGATGATATCTGTGGAAGAAACCTTACAGCAAAATCATC -AAG 154
Db 21 AspProThrLeuAspArgHisThrGlnSerThrLysLeuTyrGlyThrGlyLysGlnTyrLys 40
QY 155 GAAGCAATATGAGGAAGTACCAGGCGCTGCATCTGGGAAAAAAATCTAAATTTGTGATG 214
Db 41 GluLysAsnGlnGluAlaValArgArgLeuIleTyrGluLysAsnLeuLysPheValMet 60
QY 215 CTTCACAATCTGGACACATCAATGGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuGlyMetAsnHisLeu 80
QY 275 GGAAGCATGATGCTGTCAGAAAGTATGCTTGTGATGGGCTTCCTGAGATGCCAGCA 334
Db 81 GlyAspMetThrSerGlnGluValMetSerLeuThrSerSerLeuArgValProSerGln 100
QY 335 TGCGACAGAAATGTCATATAGTCAACTGATATGCAAAATTTGCGCATTCCTGTCAC 394
Db 101 TrpGlnArgAsnLleThrTyrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
QY 395 TGCGACAGAAAGGCTGCTGCTACTGCAAGTGAATACAGCGCTTCTGCTGCTGCTGCTG 454
Db 121 TrpArgGluLysGlyValThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
QY 455 GCTTTAGCGGCTGTGGGGCCCTGGAAAGCACAACCTGAAAGCTTAAAGCAAGAAAGCTGTG 514
Db 141 AlaPheSerAlaValGlyAlaLeuGlnAlaGlnIleuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTCTGAGTGCACAGAACTTGTAGTGTGCTCACTGCAAGAAATATGGAATTAAGGCTGC 574
Db 161 ThrLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlyAsnLysGlyLys 180
QY 575 AATGGCGGTTTCATGACAACTGCTTTCAGATATTTATGATTAACAGCGCATTCATCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGATGCAATGCACTCAAAAAGCGA 694
Db 201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGTCATATTTTCAAAATATGCAATTTGCTTTCATGCAATGCAATGCTTAAAGAA 754
Db 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValIleuLysGlu 240
QY 755 GCGTGGCGCAATTAAGACACGCTGCTGCTGCTATGATGCGACGCCATATTCCTTTC 814
Db 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGTACAGAAAGTGTGCTACTATGCAACCATCTGCTTACGAAATGCAATGATGATGATA 874
Db 261 LeuTyrArgSerGlyValIlyTyrGlyProSerCysThrGlnAsnValAsnHisGlyVal 280
QY 875 TTATGCTGCTGCTATGCTAACTTAATGGAAGAAAGTACGCTGCTGGAAGAAACAGCTG 934
Db 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
QY 935 GCGCTCAACTTGTGTCACCAAGCATATATTCGATGCGCAAGAAACAGTGGAAATCACTGT 994
Db 301 GlyHisAsnPheGlyGluGlnGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
QY 995 GGGATGCTAGTAAATCCCTTACCGCAAAATC 1027
Db 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 14
US-09-791-537-90091
; Sequence 90091, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
```



```
UY ATGAAATGGCTACTTGGGCTGCTTCGGTTGCTGCTATGCAAGTGGCCAAATGATAAA 95
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
UY GATCCGACCTCTGATCATCATCTGCAATCTCTGCAAGAAACCTACAGCAACAATC-AAG 154
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 AsprProThrLeuAspHisHisThrHisLeuTrpLysLysThrLysLysGlnTrpLys 40
UY GAATGATGATGACAAATGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 GlnLysAsnGlnClnClnAlaValArgArgLeuLeuTrpClnLysAsnLeuLysPheValMet 60
UY CTTCACAAATCTGCAACACCTCAATGGCAATGCAATGCAATGCAATGCAATGCAATCTG 274
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 LeuHisAsnLeuGlnHisSerMetGlnMetHisSerTrpAspLeuGlnMetAsnHisLeu 80
UY GGACACATGATGCTGCTGCAAGATGATATCTTTCATGCGCTTCCTGAGAGCTTCCAGCCAA 334
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
81 GlnAspMetThrSerClnClnValMetSerLeuThrSerSerLeuArgValProSerCln 100
UY TGACAGCAAAATGCTATGCTTATGCTCAATCTTATGCA:AAATGCTTATGCTTATGCTGAC 394
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 TrpClnArgAsnLcThrTrpLysSerAspProAsnArgLcLeuProAspSerValAsp 120
UY TGCAGACAGCAAGGCTGCTGCTTACTGACGCAATACCGAGCTTCTGCTGCTGCTGCTGCTG 454
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 TrpArgGlnLysGlnCysValThrGlnValLysTrpGlnLysSerCysGlnAlaCysTrp 140
UY GCTTTCAGCGCTGCTGCGCGCGCTGCAAGCAACGCAAGCTAAAAACAGCAAGCTGCTG 514
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 AlaPheSerAlaValGlnClnAlaLeuGlnClnAlaGlnLeuLysLeuLysThrGlnLysLeuVal 160
UY TCTCTGATGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 574
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 ThrLeuSerAlaClnAsnLeuValAspCysSerThrClnLysTrpClnLysLysGlnCys 180
UY AATGCGCGTTTTCATGCAACCTGCTTTCAGTATATATGATAAACAAGCGCATTTGATCA 634
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 AsnGlnLysPheMetThrThrAlaPheGlnTrpLcLcLeuAspAsnLysGlnLysLeuAspSer 200
UY GAACTTTCCTATTCCTTACAAAGCCATGCAAGCAAGTGCAGATATGCACTCAAAAAAGCGCA 694
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 AspAlaSerTrpTrpTrpLysAlaMetAspClnLysCysGlnTrpAspSerLysTrpArg 220
UY GCTGCTACATGTTTAAAGTATATGAACTTCCCTTTGGAGTGAAGTGCCTTAAAGGAA 754
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 AlaAlaThrCysSerLysTrpThrClnLeuProTrpClnLysArgLysAspValLeuLysGln 240
UY GCTGCTGCGCAATTAAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 AlaValAlaAsnLysGlnProValSerValGlnValAlaAspAlaArgHisProSerPhePhe 260
UY CTCTACAGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 874
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 LeuTrpArgSerGlnValTrpTrpGlnProSerCysThrClnAsnValAsnHisGlnVal 280
UY TTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 934
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
875 TrpValValGlnTrpClnLysAspLeuAsnGlnLysGlnTrpClnValLysAsnSerTrp 300
UY GCTCTCAACTTTCGTCAGCAAGCATATATGCGATGCGCAAGAAACAGTGCAGAAATCACTGCT 994
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GlnHisAsnPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
UY GGCAATGCTAGTATGCTCTTACCGCAAAATC 1027
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 GlnLysLeuSerPheProSerTrpProGlnLc 331
```

Search completed: June 27, 2003, 11:45:29
Job time : 296.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:29:39 ; Search time 133 Seconds
(Without alignments)
4393.190 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1997

Sequence: 1 gcatatagatagagagca.....aataatgtagatgacat 1126

Scoring table:

	BL0SUM62	Xgapop 10.0	Xgapext 0.5	Ygapop 10.0	Ygapext 0.5	Fgapop 6.0	Fgapext 7.0	Delop 6.0	Delext 7.0
--	----------	-------------	-------------	-------------	-------------	------------	-------------	-----------	------------

Searched: 1184592 seqs, 259455700 residues

Total number of hits satisfying chosen parameters: 2369184

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame.n2p.model -DBV=xip

-O/cgn2_1/USPTO.spool/US10010580/runat_27062003.104538.11821/app_query.fasta_1.1287

-DB-pending_Patents_AA_New -QFMT-fastan -SUFFIX-n2p.rapn -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-biosum62

-TRANS-human40.cdi -List=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptio -NOR=ext -HEARSTIE=500 -MINLEN=0

-MAXLEN=200000000 -USER=US10010580.ecgn_1_1_63_etunat_27062003.104538.11821

-NCPU=6 -ICPU=3 -NO_MMAP -LAKEOUBRY -NPG SCORES=0 -MAIT -DSPBLOCK=100

-LONGLOG -DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FSAPOP=6 -FSAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Pending Patents AA_New*

2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*

3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*

4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*

5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*

6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*

7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1581.5	79.2	353	6	US-10-227-425-1771
2	1581.5	79.2	353	6	US-10-227-425-2997
3	1571	78.7	331	5	US-09-724-676-81732
4	1571	78.7	331	5	US-09-724-676-81733
5	1571	78.7	331	5	US-09-724-676-81733
6	1571	78.7	331	5	US-09-724-676-81733
7	1571	78.7	331	5	US-09-724-676-81733
8	1571	78.7	331	5	US-60-452-680-23324
9	1571	78.7	331	7	US-60-453-135-14611
10	1571	78.7	331	7	US-60-453-050-14611
11	1571	78.7	331	7	US-60-455-444-7889

12	1571	78.7	331	7	US-60-466-412-14611	Sequence 14611, A
13	1568	78.5	331	6	US-10-219-051B-2757	Sequence 2757, Ap
14	1568	78.5	331	6	US-10-219-051B-10849	Sequence 10849, A
15	1565	78.4	331	6	US-10-273-577-1	Sequence 1, Appl
16	1563	78.3	331	6	US-10-273-577-6	Sequence 6, Appl
17	1560	78.1	331	6	US-10-318-584-4	Sequence 4, Appl
18	1555	77.9	331	6	US-10-273-577-8	Sequence 8, Appl
19	1553	77.8	331	6	US-10-273-577-7	Sequence 7, Appl
20	1334.5	66.8	330	6	US-10-219-051B-2755	Sequence 2755, Ap
21	1334.5	66.8	330	6	US-10-219-051B-10847	Sequence 10847, A
22	1085	54.3	217	6	US-10-273-577-5	Sequence 5, Appl
23	1083	54.2	217	6	US-10-273-577-2	Sequence 2, Appl
24	1075	53.8	217	6	US-10-273-577-4	Sequence 4, Appl
25	1073	53.7	217	6	US-10-273-577-3	Sequence 3, Appl
26	944.5	47.3	329	6	US-10-318-584-3	Sequence 3, Appl
27	939.5	47.0	329	6	US-10-318-584-2	Sequence 2, Appl
28	939.5	47.0	329	6	US-10-219-051B-6306	Sequence 6306, Ap
29	939.5	47.0	329	6	US-10-219-051B-11531	Sequence 11531, A
30	939.5	46.9	329	6	US-60-474-733-201	Sequence 201, App
31	937.5	46.9	329	6	US-10-219-051B-6304	Sequence 6304, Ap
32	937.5	46.9	329	6	US-10-219-051B-11529	Sequence 11529, A
33	934.5	46.8	329	7	US-60-452-680-23323	Sequence 23323, A
34	934.5	46.8	329	7	US-60-453-135-14610	Sequence 14610, A
35	934.5	46.8	329	7	US-60-453-050-14610	Sequence 14610, A
36	934.5	46.8	329	7	US-60-455-444-7888	Sequence 7888, Ap
37	934.5	46.8	329	7	US-60-466-412-14610	Sequence 14610, A
38	934.5	46.8	329	7	US-60-466-412-14610	Sequence 14610, A
39	926.5	46.4	329	7	US-10-405-027-3738	Sequence 3738, Ap
40	875	43.8	365	6	US-10-131-819A-12	Sequence 12, Appl
41	843	42.2	334	6	US-10-131-819A-12	Sequence 12, Appl
42	843	42.2	334	6	US-10-131-823A-12	Sequence 12, Appl
43	843	42.2	334	6	US-10-131-824A-12	Sequence 12, Appl
44	843	42.2	334	6	US-10-131-824A-12	Sequence 12, Appl
45	843	42.2	334	6	US-10-131-826A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-227-425-1771
Sequence 1771, Application US/10227425
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P1249CIN
CURRENT APPLICATION NUMBER: US/10/227,425
PRIOR FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/760,475
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 4122
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1771
LENGTH: 353
TYPE: PRT

```

ORGANISM: Homo sapiens
US-10-227-425-1771
Alignment Scores:
  Score: 1,04e-151
  Percent Similarity: 92.69%
  Host Local Similarity: 86.26%
  Query Match: 79.19%
  DB: 6
  Gaps: 1
US-10-010-580-1 (1-1126) x US-10-227-425-1771 (1-353)
UY 6 ATTAGTATGGAGCAGCAGCTTCTTC--ATGAATGGCTAGTGGCTGCTTCG 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12 LeuThrValGlyAlaThrAlaGlySerIleThrMetLysArgLeuValCysValLeu 31
UY 63 TTGTGCTCTATGACAGTGGCAGCAAGTACATTAAGATCCCACTGCAATCTGCAAT 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42 ValCysSerAlaValAlaIleuHisLysAspProThrLeuAspHisIleThrHis 51
UY 123 CTGTGAAGAAACCTACAGCAAAACATC-AAGGAAGAAATGAGAGTACGAGCGGT 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 LeuThrLysLysThrTyrGlyGlyGlnTyrLysGlnLysGlnLysGlnLysGlnLys 71
UY 182 CTATCTGGCAAAAAATCTAAATTTGTGATGCTTGCATCTGCAACCTGCAATGCA 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 LeuLeuThrProLysAsnLeuLysPheValMetLeuHisAsnLeuGluHisSerMetGly 91
UY 242 ATGCATTCATATGCTTACGATGCAACCATGCGAGACATGCTGCTGCTGCTGCTGCT 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 MetHisSerTyrAspLeuGlnLysMetAsnHisLeuGlnLysMetHisSerGlnLysValMet 111
UY 302 TTTTGTAGGTTTCTGAGAGTTCAGCAACATGCGAGAAATGCTACCTTATAGTCA 361
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 SerLeuMetSerSerLeuAlaGlyAlaProSerGlnThrPheAlaGlnHisIleThrTyrLysSer 131
UY 362 AACTCTATGCAAAATTTGCTGATTCCTGACCTGCGAGACAGCAAGCGCTGTACTGAA 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 AsnProAsnArgLysLeuProAspSerValAspThrArgGlnLysGlyCysValThrGln 151
UY 422 GTCAATATACAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 ValLysTyrGlnLysSerGlyAlaCysThrAlaHisSerAlaValGlyAlaLeuGln 171
UY 482 GTATCACTGAATTTAAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 AlaGlnLeuLysLeuLysThrGlyLysLeuValSerLeuSerAlaGlnAsnLeuValAsp 191
UY 542 TGTCACTGCAAAATTTGCGAATTAAGCGCTGCGAATGCGCTTTCATGACAGCTGCTTC 601
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 CysSerThrGlnLysTyrGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 211
UY 602 CACTATATATATGCAACAGCGCATGATGATGCAACAGCTTTCATGCTTACAAAGCGCAT 661
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 GlnTyrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 231
UY 662 AATGGAAGAAATTTAAATAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 721
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 AspLysCysGlnTyrAspSerLysTyrAlaAlaIleThrCysSerLysTyrThrGln 251
UY 722 CTTCCTTGGCAGCATGCGCTTAAAGAGCGCTGCGCAATTAAGAGCGCTGCTTC 781
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 LeuProTyrGlnLysGlnLysPheValLeuLysGlnLysAlaValAlaAsnLysGlnProValSer 271
UY 782 TGTGCTCTATGACAGTGGCAGCAAGTACATTAAGATCCCACTGATCATCTGCAAT 841
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 ValCysSerAlaValAlaIleuHisLysAspProThrLeuAspHisIleThrHis 291
UY 842 CTATCTGTATTAAGATGTAATATGAGTATATATGCTTATGCTTATGCTTATGCTTAT 901
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 ProSerCysThrGlnAsnValAsnHisGlnValLysValLysValLysValLysValLys 311

```

```

UY 902 CGGAAGAGTACAGCTTGTGCAAAAAACAGCTGCGCGCTCACTTGTGGACCAAGCATAT 961
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 GlyLysGlnTyrThrLeuValLysAsnSerTrpGlnHisAsnPheGlyGlnLysLys 331
UY 962 ATTGGATGGCAAGAAAGAGTGAATATCTGTGGAATGTTATTTATCTGTTATGCA 1021
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 IleArgMetAlaArgAsnLysLysAsnHisCysGlyIleAlaSerPheProSerTyrPro 351
UY 1022 GAATC 1027
   |||||
Db 352 GluIle 353
RESULT 2
US-10-227-425-2997
; Sequence 2997, Application US/10227425
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT249C1N
; CURRENT APPLICATION NUMBER: US/10/227,425
; PRIORITY FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/760,475
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO: 2997
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-227-425-2997
Alignment Scores:
  Pred. No.: 1,04e-151
  Score: 1581.50
  Percent Similarity: 92.69%
  Best Local Similarity: 86.26%
  Query Match: 79.19%
  DB: 6
  Gaps: 1
US-10-010-580-1 (1-1126) x US-10-227-425-2997 (1-353)
UY 6 ATTAGTATGGAGCAGCAGCTTCTTC--ATGAATGGCTAGTGGCTGCTTCG 62
   ..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12 LeuThrValGlyAlaThrAlaGlySerIleThrMetLysArgLeuValCysValLeu 31
UY 63 TTGTGCTCTATGACAGTGGCAGCAAGTACATTAAGATCCCACTGATCATCTGCAAT 122
   ..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 ValCysSerAlaValAlaIleuHisLysAspProThrLeuAspHisIleThrHis 51
UY 123 CTGTGAAGAAACCTACAGCAAAACATC-AAGGAAGAAATGAGAGTACGAGCGGT 181
   ..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 LeuThrLysLysThrTyrGlyGlyGlnTyrLysGlnLysGlnLysGlnLysGlnLys 71
UY 182 CTATCTGGCAAAAAATCTAAATTTGTGATGCTTGCATCTGCAACCTGCAATGCA 241

```

```

Db      72  LeuiletrpclulysasnleuylsphevalmetleuHisasnleuglHisserMetgly 91
QY      242 ATGCATTATATATCTAGCATGCAACCATCTGGAGACATGCTGTGACAGAGATGATA 301
Db      92  MetHisserlyrAspleuglyMetasnHisleuglyAspMetThrsergluglyValmet 111
QY      302 TCTTTGATGGGTCCCTGAGAGTCCAGCCCAATGGAGAAATGCAATCTTATAGTCA 361
Db      112 SerleuMetSerSerleuArGvalProserGlnThrPalaArgasnleuThrlyrLysSer 131
QY      362 AACTCTAATAGAAATTCCTGATTTCTGTGAGTGGAGAGAGAAAGGCTGTGTACTGAA 421
Db      132 AspProasnArGleleuProAspSerValAspTrpArgIulysglYcysValThrGlu 151
QY      422 GTCAAAATACCAAGGCTTCTGTGGCTGTGTGGCTTTCAGCCCTGGGGGCCCTGGAA 481
Db      152 ValllystyrGlnleuLysersglYalaCysTrpAlaPheSerAlaValGlyAlaLeuGlu 171
QY      482 GCACAACTGAAGCTAAACAGAGAAAGCTGTCTCTAGTGCACAGAACTTGTAGAT 541
Db      172 AlaGlnleuLysleuLysThrGlyLysleuValSerleuSerAlaGlnAsnleuValasp 191
QY      542 TGCTCAACTGAAAAATATGGGAATTAAGGCTGCAATGGGGGTTTCATGACACTGCTTC 601
Db      192 CysSerThrGluLyslyrGlyAsnLysglYcysAsnGlyLysPheMetThrThrAlaPhe 211
QY      602 CAGTATATTATGATTAACAACGGCATTCATGACAGCTTCATCCCTACAAAGCCATG 661
Db      212 GlnIyrlleleAspAsnLysglYleAspSerAlaSerlyrProIyrlYsAlaMet 231
QY      662 AATGGAAGTGCAGATATGACTCAAAAAGCGAGCTGCCACATGTTCAAGTACTGAA 721
Db      232 AspGlnLysCysGlnIyrlYrAspSerLysTyArGAlaAlaThrCysSerlyrThrGlu 251
QY      722 CTTCCTCTTGGCAGTGAAGATGCTTAAGAGCTGTGGCCATTAAGAGACTGTGCT 781
Db      252 LeuProIyrlYrArgIuAspValleuLysglYAlaValAlaAsnLysglYrProValSer 271
QY      782 GTGGCTATAGATCGCAGCCACTATTTCTTCTCTGTACAGAAAGTGTGTACTGAA 841
Db      272 ValGlyValAspAlaArGHisProSerPhePheLeuTyArGSerGlyValTyrglu 291
QY      842 CCATCTGACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
Db      292 ProserCysThrGlnAsnValAsnHisGlyValleuValValGlyTyrglyAspleuAsn 311
QY      902 GGGAAAGACTACTGGCTTGTGAAAAACAGCTGGGCTCAACTTGTGACCAAGGATAT 961
Db      312 GlyLysgluTyrlYrPleuValLysAsnSerTrpGlyHisAsnPhelGlyGluGluGlyr 331
QY      962 ATTGGGATGGCAGAAACAGTGAATCACTGTGGGATTCCTAGTTATCCCTTTACCA 1021
Db      332 IleArMetAlaArGAsnLysglYAsnHisCysGlyleAlaSerPheProSerTyPro 351
QY      1022 GAATATC 1027
Db      352 GluIle 353

```

```

; ORGANISM: Homo sapiens
US-09-724-676-81732

Alignment Scores:
Pred. No.: 1,19e-150
Score: 1571.00
Percent Similarity: 93.35%
Best Local Similarity: 87.61%
Query Match: 78.67%
DB: 5
Gaps: 0

US-10-010-580-1 (1-1126) x US-09-724-676-81732 (1-331)

QY      36  ATGAAGATGGCTGTGGGCTGCTCCGTTGCTGCTCTATGACAGTGCACAAATGACATAA 95
Db      1  MetLysArGleuValCysValleuLeuValCysSerSerAlaValAlaGlnleuHisLys 20
QY      96  GATCCCACTCTGATCATCATCTGAAATCTCTGGAAGAAACCTACAGCAACATC- AAG 154
Db      21  AspProThrLeuAspHisThrPheLysleuTrpLysLysThrTyrglyLysGlnTyrlYs 40
QY      155 GAAGCAATAGACAACTACACAGCGCTCATCTGGGAAAAAAATCTAAATTTGTGATG 214
Db      41  GluLysAsnGluGluAlaValArGArGleuIleTrpGluLysAsnleuLysPheValMet 60
QY      215 CTTCACAACTCTGGAACACTCAATGGGAATGCATTCATATGATCTAGGCATGACCATGTG 274
Db      61  LeuHisAsnleuGlnHisSerMetCylmethHisSerlyrArPleuGlyMetAsnHisLys 80
QY      275 GGAGCATGATCATCTGGTGAAGAACTGATATCTTTGATGGCTCCCTGAGATCCCAACCA 334
Db      81  GlyAspMetThrSerGluGluValMetSerLeuMetSerSerLeuArGValProserGln 100
QY      335 TGGCAGACAAATGTACTTATAGTCAACACCTATATGCAAAATTCCTGATTCGTGTGAC 394
Db      101 TrpGlnArGAsnIleThrTyrlYrLysSerAsnProAsnArGlyleuProAspSerValasp 120
QY      395 TGGAGAGAGAGAGGCTGTGTTACTGAACTGAATATCCAGGGCTTCTGTGCTTGTGG 454
Db      121 TrpArGluLysglYcysValThrGluValLysTyrglnGlySerCysGlyAlaCysTrp 140
QY      455 GCTTCAGAGGCTGTGGGGCCCTGGAGACATCAACTGGAAGTGAATTAAGTGAAGTGTG 514
Db      141 AlaPheSerAlaValGlyAlaLeuGluAlaGlnleuLysleuLysThrGlyLysleuVal 160
QY      515 TCTGTAGTGCACAGAACTTGTAGATTCCTCAACTGCAAAAATATGGCAATTAAGCGCTGC 574
Db      161 SerleuSerAlaGlnAsnleuValAspCysSerThrGluLysTyrglyAsnLysglYcys 180
QY      575 AATGGCGGTTTGTATGACAACTGCTTTCGATGATATATTATGATTAACAGGCAATGATTC 634
Db      181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrlleleAspAsnLysglYleAspSer 200
QY      635 GAAGCTTCCTATCCCTACAAAGCCATGAATGGAAGTGCATATGATGATTAAGAAAGCA 694
Db      201 AspAlaSerTyrlYrProTyrlYsAlaMetAsnGlnLysCysGlnTyrlYrAspSerLysTyArG 220
QY      695 GCTGCCAATGTTCAAAGTATAGTGAACCTTCCCTTGGCAGTGAAGATCCCTTAAGAA 754
Db      221 AlaAlaThrCysSerLysTyrlYrThrGlnleuProTyrlYrArGValAspValleuLysGlu 240
QY      755 GCTGGGCAATTAAGAGCTGTGTGTGATTAAGTGAATGATTAAGTGAATTAAGTGA 814
Db      241 AlaValAlaAsnLysglYrProValSerValGlyValAspAlaArGHisProSerPhe 260
QY      815 CTGTACAGAAAGTGTGTACTATGAACTGATCACTGATCTGAGAAATGTAATCATGAGTA 874
Db      261 LeuTyrlYrArGSerGlyValTyrlYrGluProserCysThrGlnAsnValAsnHisGlyVal 280
QY      875 TTATGCTGTGGTATGATGATCACTTAATGCAAGAAAGCTACTGCTTGTGAAAAACAGCTGC 934
Db      281 LeuValValGlyTyrglyAspleuAsnGlyLysGlnTyrlYrPleuValLysAsnSerTrp 300

```

```

RESULT 3
US-09-724-676-81732
; Sequence 81732, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 81732
; LENGTH: 331
; TYPE: PRK

```

QY 945 GGCCTCACTTGGTACATAGATATATGATGAGAAACAGTGGAAATCATCTG 994
DB 301 G1YH1ASnphc1uylugluclyltyr11leatgmet1a1atgasnlysg1yasn1iscys 320
QY 995 GGCATTGCTAGTATTCCTCTTACCCAGAAATC 1027
DB 321 G1YH1eAlaserpheroSertyrProguille 331

RESULT 4
US-09-724-676-81733
: Sequence 81733, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 81733
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-81733

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
DB: 5 Gaps: 0

US-10-010-580-1 (1-1126) x US-09-724-676-81733 (1-331)

QY 36 ATGAAATGCTAGTGGTGTCTGTCTGTCTATGAGTGGACAGTACATATAA 95
DB 1 MetLysArgLeuValAlcysValLeuValCysSerSer1aValAlaGlnLeuHisLys 20
QY 96 GATCCCACTGCTGATCATCTGATGCAAGAAACCTACAGCAAAATC-AAG 154
DB 21 AspProthLeuAspHisHis1stPrlHisLeuTrpLysLysThrTrpGlyGlnTyrLys 40
QY 155 GAAGAGATGAGAAATGATACGCTGTCTATGTTGAAAAAATCTAAATTTGATG 214
DB 41 GlutylAsnGlnGluAlaValArgArgLeuLeuTrpGlnLysAsnLeuLysPheValMet 60
QY 215 CTTCACAAATCTGAACATCATGAGAAATGATATATGATCATGSCATGAACCATCTG 274
DB 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlyMetAsnHisLeu 80
QY 275 GCAACATGACCTGGTGAAGATGATATCTTGTGATGGCTCCCTGACACTTCCACGCCA 334
DB 81 G1YASpMetThrSerGluGluValMetSerLeuMetSerSerLeuArgValProSerGln 100
QY 335 TGACAGGAAATGCTATATAGTAAATCTAAATGAGAAATGCTGATTTGTGGAC 394
DB 101 TrpGlnArgAsn1LeuThrTrpLysSerAsnProAsnArg11LeuProAspSerValAsp 120
QY 395 TGACAGGAAAGGCTGTGTTACTGAAGTAAATCTACAGGTTCTGTGGTGG 454
DB 121 TrpArgGluLysGlyCysVal1TrpGluValLysTrpGlnLysSerGlyAlaCysTrp 140
QY 455 GATTTCAGGCTGTGGGCGCCCTGGAGACAGCACTGAAGCTAAAGACAGCAAGCTGGTG 514
DB 141 AlaPheSerAlaValAlcylAlaLeuGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTCTGATGTAAGAAATTTAGATTTCTCAATGAAATATGGGAATAAAGCTGC 574
DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrClnLysTrpGlyAsnLysGlyCys 180
QY 575 AATATGATTTTATATGAAATTTTATATATATTATATACAAAGCATTTGATTA 634

DB 181 AsnGlyLysPheMetThrThrAlaPheGlnTyrLle11eAspAsnLysGlyLleAspSer 200
QY 635 GAAGCTTCTATCCCTACAAAGCCATGATGGAAGTGCACATATGACTCAAAAAAGCGA 694
DB 201 AspAlaSerTyrTrpTrpLysAlaMetAspClnLysCysGlnTyrAspSerLysTrpArg 220
QY 695 GCGCCACATGTTCAAGATATACAGTGAATCTTGGACAGTGAAGATGCTTAAAGAA 754
DB 221 AlaAlaThrCysSerLysTrpThrGluLeuProTyrGlyArgGlnAspValLeuLysGln 240
QY 755 GCTGTGCGCAATAAAGACCTGTGTCTGTGCTATAGATGCGAGCCACTATCTTCTTC 814
DB 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhe 260
QY 815 CTTGACAGAGTGGCTGTCTATGATACCATGCTGTGCTATGATGCGAGCCACTATCTTCTTC 874
DB 261 LeuTyrArgSerGlyAlaTyrTrpGluProSerCysThrGlnAsnValAsnHisGlyAl 280
QY 875 TTAGTGTGCTATGCTAACCTTAATGGAAAGACTAGCTGTGAAAAACAGCTGG 934
DB 281 LeuValValGlyTyrGlyLysPheLysGlnLysGlyLysGlyLysTrpLeuValLysAsnSerTrp 300
QY 935 GGCCTCACTTGGTGTGACCAAGGATATATGCGATGCGCAAGCAAGTGAATCATCTGT 994
DB 301 G1YH1ASnphc1uylugluclyltyr11leatgmet1a1atgasnlysg1yasn1iscys 320
QY 995 GGCATTGCTAGTATTCCTCTTACCCAGAAATC 1027
DB 321 G1YH1eAlaserpheroSertyrProguille 331

RESULT 5
US-09-724-676A-81732
: Sequence 81732, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 81732
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-81732

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
DB: 5 Gaps: 0

US-10-010-580-1 (1-1126) x US-09-724-676A-81732 (1-331)

QY 36 ATGAAATGCTAGTGGTGTCTGTCTGTCTATGAGTGGACAGTACATATAA 95
DB 1 MetLysArgLeuValAlcysValLeuValCysSerSer1aValAlaGlnLeuHisLys 20
QY 96 GATCCCACTGCTGATCATCTGATGCAAGAAACCTACAGCAAAATC-AAG 154
DB 21 AspProthLeuAspHisHis1stPrlHisLeuTrpLysLysThrTrpGlyGlnTyrLys 40
QY 155 GAAGAGATGAGAAATGATACGCTGTCTATGTTGAAAAAATCTAAATTTGATG 214
DB 41 GlutylAsnGlnGluAlaValArgArgLeuLeuTrpGlnLysAsnLeuLysPheValMet 60
QY 215 CTTCACAAATCTGAACATCATGAGAAATGATATATGATCATGSCATGAACCATCTG 274
DB 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlyMetAsnHisLeu 80


```

RESULT 7
US-60 452-680-23124
Sequence 23124, Application US/60452680
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
APPLICANT: GROPE, Andrew
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01450
CURRENT APPLICATION NUMBER: US/60/452,680
CURRENT FILING DATE: 2003-03-07
NUMBER OF SEQ ID NOS: 116213
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23124
LENGTH: 331
TYPE: PR1
ORGANISM: Homo sapiens
US-60 452-680-23124

```



```

OY 335 TGGCAGAAATGCTATTAGTCAACTCTAATGAAATGGCTGATCTGTGGC 394
DB 101 TTPGlnATgAsnIleHrTyrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
OY 395 TGGCAGAGAAAGGCGTGTCTTACTGAACTGAAATACAGAGGCTTGTGTGCTGTGGTGG 454
DB 121 TTPArgGluLysGlyCysValThrGluValIleLysTyrGlnGlySerCysGlyAlaCysTrp 140
OY 455 GCTTTCACGCGTGTGGGGCGCCCTGGAACGACACGTAAGCTTAAAAACGAGAACTGTGC 514
DB 141 AlaPheSerAlaValAlaGluLeuIleAsnLeuLysLeuLysThrGlyLysLeuVal 160
OY 515 TCTGTGAGTCACAGAACTGTGATGCTCAACTGAAATATGGGAAATGAAGGCTGC 574
DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrGlyAsnLysClyCys 180
OY 575 AATGGCGGTTTCATGACACAGCTGTTCACATATATATGATTAACACGCGATTGATTC 634
DB 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGATGAAAGATGACATATGACTCAAAAAAGCGA 694
DB 201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
OY 695 GCTGGCAGATGTTCAAGATATACGAACTTCCCTTGGCAGCTGAAGTGCCTTAAGAA 754
DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
OY 755 GCTGGCGCAATTAAGACGCTGTCTGTGCTATAGTCGAGCCGCTATTCCTTCTTC 814
DB 241 AlaValAlaAsnLysGlyPheValSerValGlyValAlaAspAlaGlnHisProSerPhePhe 260
OY 815 CTGTACAGAACTGCTGTCTACTATGAAACCATCTGTACTGACAGATGTAATCATAGAGTA 874
DB 261 LeuTyrArgSerGlyValTyrTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
OY 875 TTATGGTGGCTATGTAACCTTAATGGGAAAGACTGAGCTTGGGAAACAGCTGG 934
DB 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
OY 935 GGCTCAACTTTGGTACCAAGATATATTCGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 301 GlyHisAsnPheGlyGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
OY 995 GCGATGCTAGTTATCCCTCTTACCAGAAATC 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 12
US-60-466-412-14611
: Sequence 14611, Application US/60466412
: GENERAL INFORMATION:
: APPLICANT: CARBIL, Michele
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: CLO001466
: CURRENT APPLICATION NUMBER: US/60/466,412
: CURRENT FILING DATE: 2003-04-30
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14611
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-466-412-14611

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1

```

```

DB: 7 Gaps: 0
US-10-010-580-1 (1-1126) x US-60-466-412-14611 (1-331)
OY 36 ATGAATGAGTACGTAGTGGCGCTGCTTGGCTTATCCAGTGGCAGACATACATAAA 95
DB 1 MetLysArgLeuValCysValLeuLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
OY 96 GATCCCACTGTGATCATCACTGAAATCTCTGGAGAAAAACCTACAGCAAACTC -AAG 154
DB 21 AspProThrLeuAsnPhnHisIstRphHisLeuTrpLysLysThrTyrGlyLysGlnTyrLys 40
OY 155 GAAGAGAAATGAGAAAGTAGCAGCGGCTTCATCTGGGAAAAAATTTAAATTTGTATG 214
DB 41 GluLysAsnGluAlaValAlaArgArgLeuIleTTPGlnLysAsnLeuLysPheValMet 60
OY 215 CTTTCACAATCTGGAGACCTCAATGGAATTCATTCATATGATCAGTGAACCATTCG 274
DB 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuGlyMetAsnHisLeu 80
OY 275 GGAAGATGACCTGTGTAAGAGATGATATCTTATGAGGTTCCCTGAGAGCTCCACCAA 334
DB 81 GlyAspMetThrSerGluGluValaMetSerLeuMetSerSerLeuArgValProSerGln 100
OY 335 TGGCAGAGAAATGCTCACTTATAGTCAAACTTAATCAGAAATGGCTGATTCGTGAGC 394
DB 101 TTPGlnArgAsnIleHrTyrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
OY 395 TGGCAGAGAAAGGCGTGTCTTACTGAACTGAAATACAGAGGCTTGTGTGCTGTGGTGG 454
DB 121 TTPArgGluLysGlyCysValThrGluValIleLysTyrGlnGlySerCysGlyAlaCysTrp 140
OY 455 GCTTTCACGCGTGTGGGGCGCCCTGGAACGACACGTAAGCTTAAAAACGAGAACTGTGC 514
DB 141 AlaPheSerAlaValAlaGluLeuIleAsnLeuLysLeuLysThrGlyLysLeuVal 160
OY 515 TCTGTGAGTCACAGAACTGTGATGCTCAACTGAAATATGGGAAATGAAGGCTGC 574
DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrClyAsnLysClyCys 180
OY 575 AATGGCGGTTTCATGACACGCTGTTCACATATATGATTAACACGCGATTGATTC 634
DB 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGATGAAAGATGACATATGACTCAAAAAAGCGA 694
DB 201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
OY 695 GCTGGCAGATGTTCAAGATATACGAACTTCCCTTGGCAGCTGAAGTGCCTTAAGAA 754
DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
OY 755 GCTGGCGCAATTAAGACGCTGTCTGTGCTATAGTCGAGCCGCTATTCCTTCTTC 814
DB 241 AlaValAlaAsnLysGlyPheValSerValGlyValAlaAspAlaGlnHisProSerPhePhe 260
OY 815 CTGTACAGAACTGCTGTCTACTATGAAACCATCTGTACTGACAGATGTAATCATAGAGTA 874
DB 261 LeuTyrArgSerGlyValTyrTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
OY 875 TTATGGTGGCTATGTAACCTTAATGGGAAAGACTGAGCTTGGGAAACAGCTGG 934
DB 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
OY 935 GGCTCAACTTTGGTACCAAGATATATTCGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 301 GlyHisAsnPheGlyGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
OY 995 GCGATGCTAGTTATCCCTCTTACCAGAAATC 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 13

```



```
DB 261 leuYTAqSeGlyValTtYrGluProSeRCysThrGlnAsnValAsnHisGlyVal 280
GY 875 TTATGTTTGGTTATGTTATGCTTAATGGAAGACTACTGCTTGTGAAAAACAGCTGG 934
DB 281 leuValValGlyTyrGlyAspIleuAsnGlyLysGluTyrTrpIleuValLysAsnSerTrp 300
GY 945 GGCCTCAACTTTGGTGACCAAGCATATATCGATGGCAAGAAACACCTGCAATCAGCTGT 994
DB 301 GlyHisAsnPhenGlyGluGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
GY 995 GGCATTCCTACTTATCCCTCTTACCCAGCAATC 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331
```

Search completed: June 27, 2003, 11:50:01
Job time : 141 secs